DOCUMENT RESUME

ED 474 612 UD 035 559

AUTHOR	Mac Iver, Martha Abele; Kemper, Elizabeth; Stringfield, Sam
TITLE	The Baltimore Curriculum Project: Final Report of the Four-Year Evaluation Study.
INSTITUTION	Center for Research on the Education of Students Placed At Risk, Baltimore, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.; Abell Foundation, Baltimore, MD.
REPORT NO	CRESPAR-RN-62
PUB DATE	2003-02-00
NOTE	45p.
CONTRACT	R-117-D40005
AVAILABLE FROM	Johns Hopkins University, Center for Research on the Education of Students Placed At Risk, 3003 North Charles Street, Suite 200, Baltimore, MD 21218. Tel: 410-516-8800; Fax: 410-516-8890; Web site: http://www.csos.jhu.edu/crespar/reports.htm.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Educational Change; Elementary Education; Grade 2; Kindergarten; *Mathematics Achievement; Program Evaluation; Public Schools; *Reading Achievement; Teaching Methods; *Urban Schools
IDENTIFIERS	Baltimore City Public Schools MD; *Core Knowledge; *Direct Instruction

ABSTRACT

This 4-year study examined implementation of the Baltimore Curriculum Project (BCP) in six Baltimore City public schools. BCP used a combination of direct instruction (DI) and core knowledge as its reform curriculum. Each school was demographically matched with a similar, within-district school. Two cohorts of students were followed throughout the 4 years (students who were in either kindergarten or grade 2 during 1996-97). Interviews with principals and DI coordinators and focus groups with teachers were conducted each year to gauge staff perceptions of the innovation. In the first 3 years, classroom observations were made in BCP schools. Overall, DI curriculum and instructional methods were implemented in BCP schools, though implementation did not proceed at the desired rate in kindergarten until year 4. Implementation of core knowledge was not envisioned to begin until year 3 and proceeded more slowly than DI implementation. Teachers expressed positive views of both DI and core knowledge, though they had some frustrations. Achievement tests data indicated mixed results for students, depending on subject, grade level, and school. Results were most positive for mathematics computation. DI students made the most significant improvements in mathematics computation and reading. An appendix includes a comparison of BCP and control schools. (Contains 50 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 474 612

UD UDTR 0052

CRESPAR

THE BALTIMORE CURRICULUM PROJECT

Final Report of the Four-Year Evaluation Study

Martha Abele Mac Iver
Elizabeth Kemper
Sam Stringfield

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Report No. 62 / February 2003

FUNDED BY
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT
U.S. DEPARTMENT OF EDUCATION

JOHNS HOPKINS UNIVERSITY & HOWARD UNIVERSITY

CENTER FOR RESEARCH ON THE EDUCATION OF STUDENTS PLACED AT RISK

UD 035 559

BEST COPY AVAILABLE

THE BALTIMORE CURRICULUM PROJECT

Final Report of the Four-Year Evaluation Study

Martha Abele Mac Iver

Elizabeth Kemper

Sam Stringfield

Report No. 62

February 2003

This report was published by the Center for Research on the Education of Students Placed At Risk (CRESPAR), a national research and development center supported by a grant (No. R-117-40005) from the Office of Educational Research and Improvement (OERI), U.S. Department of Education. The conduct of much of the research in this report was supported by The Abell Foundation of Baltimore, Maryland. The content or opinions expressed herein are those of the authors, and do not necessarily reflect the views of the Department of Education, any other agency of the U.S. Government, or other funders. Reports are available from: Publications Department, CRESPAR/Johns Hopkins University; 3003 N. Charles Street, Suite 200, Baltimore MD 21218. An on-line version of this report is available at our web site: *www.csos.jhu.edu*.

Copyright 2003, The Johns Hopkins University, all rights reserved.

CRESPAR TECHNICAL REPORTS

1. *The Talent Development High School: Essential Components*—V. LaPoint, W. Jordan, J.M. McPartland, D.P. Towns
2. *The Talent Development High School: Early Evidence of Impact on School Climate, Attendance, and Student Promotion*—J.M. McPartland, N. Legters, W. Jordan, E.L. McDill
3. *The Talent Development Middle School: Essential Components*—S. Madhere, D.J. Mac Iver
4. *The Talent Development Middle School: Creating a Motivational Climate Conducive to Talent Development in Middle Schools: Implementation and Effects of Student Team Reading*—D.J. Mac Iver, S.B. Plank
5. *Patterns of Urban Student Mobility and Local School Reform: Technical Report*—D. Kerbow
6. *Scaling Up: Lessons Learned in the Dissemination of Success for All*—R.E. Slavin, N.A. Madden
7. *School-Family-Community Partnerships and the Academic Achievement of African American, Urban Adolescents*—M.G. Sanders
8. *Asian American Students At Risk: A Literature Review*—S.-F. Siu
9. *Reducing Talent Loss: The Impact of Information, Guidance, and Actions on Postsecondary Enrollment*—S.B. Plank, W.J. Jordan
10. *Effects of Bilingual Cooperative Integrated Reading and Composition on Students Transitioning from Spanish to English Reading*—M. Calderón, R. Hertz-Lazarowitz, G. Ivory, R.E. Slavin
11. *Effective Programs for Latino Students in Elementary and Middle Schools*—O.S. Fashola, R.E. Slavin, M. Calderón, R. Durán
12. *Detracking in a Racially Mixed Urban High School*—R. Cooper
13. *Building Effective School-Family-Community Partnerships in a Large Urban School District*—M.G. Sanders
14. *Volunteer Tutoring Programs: A Review of Research on Achievement Outcomes*—B.A. Wasik
15. *Working Together to Become Proficient Readers: Early Impact of the Talent Development Middle School's Student Team Literature Program*—D.J. Mac Iver, S.B. Plank, R. Balfanz
16. *Success for All: Exploring the Technical, Normative, Political, and Socio-Cultural Dimensions of Scaling Up*—R. Cooper, R.E. Slavin, N.A. Madden
17. *MathWings: Early Indicators of Effectiveness*—N.A. Madden, R.E. Slavin, K. Simons
18. *Parental Involvement in Students' Education During Middle School and High School*—S. Catsambis, J.E. Garland
19. *Success for All/Éxito Para Todos: Effects on the Reading Achievement of Students Acquiring English*—R.E. Slavin, N.A. Madden
20. *Implementing a Highly Specified Curricular, Instructional, and Organizational School Design in a High-Poverty, Urban Elementary School: Three Year Results*—B. McHugh, S. Stringfield
21. *The Talent Development Middle School: An Elective Replacement Approach to Providing Extra Help in Math—The CATAMA Program (Computer and Team-Assisted Mathematics Acceleration)*—D.J. Mac Iver, R. Balfanz, S.B. Plank
22. *School-Family-Community Partnerships in Middle And High Schools: From Theory to Practice*—M.G. Sanders, J.L. Epstein
23. *Sources of Talent Loss Among High-Achieving Poor Students*—W.J. Jordan, S.B. Plank
24. *Review of Extended-Day and After-School Programs and Their Effectiveness*—O.S. Fashola
25. *Teachers' Appraisals of Talent Development Middle School Training, Materials, and Student Progress*—E. Useem
26. *Exploring the Dynamics of Resilience in an Elementary School*—S.M. Nettles, F.P. Robinson
27. *Expanding Knowledge of Parental Involvement in Secondary Education: Effects on High School Academic Success*—S. Catsambis
28. *Socio-Cultural and Within-School Factors That Effect the Quality of Implementation of School-Wide Programs*—R. Cooper
29. *How Students Invest Their Time Out of School: Effects on School Engagement, Perceptions of Life Chances, and Achievement*—W.J. Jordan, S.M. Nettles
30. *Disseminating Success for All: Lessons for Policy and Practice*—R.E. Slavin, N.A. Madden
31. *Small Learning Communities Meet School-to-Work: Whole-School Restructuring for Urban Comprehensive High Schools*—N.E. Legters
32. *Family Partnerships with High Schools: The Parents' Perspective*—M.G. Sanders, J.L. Epstein, L. Connors-Tadros
33. *Grade Retention: Prevalence, Timing, and Effects*—N. Karweit
34. *Preparing Educators for School-Family-Community Partnerships: Results of a National Survey of Colleges and Universities*—J.L. Epstein, M.G. Sanders, L.A. Clark
35. *How Schools Choose Externally Developed Reform Designs*—A. Datnow
36. *Roots & Wings: Effects of Whole-School Reform on Student Achievement*—R.E. Slavin, N.A. Madden
37. *Teacher Collaboration in a Restructuring Urban High School*—N.E. Legters
38. *The Child First Authority After-School Program: A Descriptive Evaluation*—O.S. Fashola
39. *MathWings: Effects on Student Mathematics Performance*—N.A. Madden, R.E. Slavin, K. Simons
40. *Core Knowledge Curriculum: Three-Year Analysis of Implementation and Effects in Five Schools*—B. McHugh, S. Stringfield
41. *Success for All/Roots & Wings: Summary of Research on Achievement Outcomes*—R.E. Slavin, N.A. Madden
42. *The Role of Cultural Factors in School Relevant Cognitive Functioning: Synthesis of Findings on Cultural Contexts, Cultural Orientations, and Individual Differences*—A.W. Boykin, C.T. Bailey
43. *The Role of Cultural Factors in School Relevant Cognitive Functioning: Description of Home Environmental Factors, Cultural Orientations, and Learning Preferences*—A.W. Boykin, C.T. Bailey
44. *Classroom Cultural Ecology: The Dynamics of Classroom Life in Schools Serving Low-Income African American Children*—C.M. Ellison, A.W. Boykin, D.P. Towns, A. Stokes
45. *An "Inside" Look at Success for All: A Qualitative Study of Implementation and Teaching and Learning*—A. Datnow, M. Castellano
46. *Lessons for Scaling Up: Evaluations of the Talent Development Middle School's Student Team Literature Program*—S.B. Plank, E. Young
47. *A Two-Way Bilingual Program: Promise, Practice, and Precautions*—M. Calderón, Argelia Carreón
48. *Four Models of School Improvement: Successes and Challenges in Reforming Low-Performing, High-Poverty Title I Schools*—G.D. Borman, L. Rachuba, A. Datnow, M. Alberg, M. Mac Iver, S. Stringfield, S. Ross
49. *National Evaluation of Core Knowledge Sequence Implementation: Final Report*—S. Stringfield, A. Datnow, G. Borman, L. Rachuba
50. *Core Knowledge Curriculum: Five-Year Analysis of Implementation and Effects in Five Maryland Schools*—M.A. Mac Iver, S. Stringfield, B. McHugh
51. *Effects of Success for All on TAAS Reading: A Texas Statewide Evaluation*—E.A. Hurley, A. Chamberlain, R.E. Slavin, N.A. Madden
52. *Academic Success Among Poor and Minority Students: An Analysis of Competing Models of School Effects*—G.D. Borman, L.T. Rachuba
53. *The Long-Te-m Effects and Cost-Effectiveness of Success for All*—G.D. Borman, G.M. Hewes
54. *Neighborhood and School Influences on the Family Life and Mathematics Performance of Eighth-Grade Students*—S. Catsambis, A.A. Beveridge
55. *The Public School Superintendency in the 21st Century: The Quest to Define Effective Leadership*—J.Y. Thomas
56. *Local School Boards Under Review: Their Role and Effectiveness in Relation to Students' Academic Achievement*—D. Land
57. *Program Development in the National Network of Partnership Schools: A Comparison of Elementary, Middle, and High Schools*—M.G. Sanders, B.S. Simon
58. *Developing Transitional Programs for English Language Learners: Contextual Factors and Effective Programming*—D. August
59. *Comprehensive School Reform and Student Achievement: A Meta-Analysis*—G.D. Borman, G.M. Hewes, L.T. Overman, S.Brown
60. *Cultural Issues Related to High School Reform: Deciphering the Case of Black Males*—W.J. Jordan, R. Cooper
61. *Supporting the Development of English Literacy in English Language Learners: Key Issues and Promising Practices*—D. August
62. *The Baltimore Curriculum Project: Final Report of the Four-Year Evaluation Study*—M. Mac Iver, E. Kemper, S. Stringfield

PLEASE NOTE:

Reports available from Publications Dept., Johns Hopkins University/CRESPAR, 3003 N. Charles Street, Suite 200, Baltimore MD 21218; 410-516-8808 phone/410-516-8890 fax; *www.csos.jhu.edu*

BEST COPY AVAILABLE

THE CENTER

Every child has the capacity to succeed in school and in life. Yet far too many children fail to meet their potential. Many students, especially those from poor and minority families, are placed at risk by school practices that sort some students into high-quality programs and other students into low-quality education. CRESPAR believes that schools must replace the "sorting paradigm" with a "talent development" model that sets high expectations for all students, and ensures that all students receive a rich and demanding curriculum with appropriate assistance and support.

The mission of the Center for Research on the Education of Students Placed At Risk (CRESPAR) is to conduct the research, development, evaluation, and dissemination needed to transform schooling for students placed at risk. The work of the Center is guided by three central themes—ensuring the success of all students at key development points, building on students' personal and cultural assets, and scaling up effective programs—and conducted through research and development programs in the areas of early and elementary studies; middle and high school studies; school, family, and community partnerships; and systemic supports for school reform, as well as a program of institutional activities.

CRESPAR is organized as a partnership of Johns Hopkins University and Howard University, and supported by the National Institute on the Education of At-Risk Students (At-Risk Institute), one of five institutes created by the Educational Research, Development, Dissemination and Improvement Act of 1994 and located within the Office of Educational Research and Improvement (OERI) at the U.S. Department of Education. The At-Risk Institute supports a range of research and development activities designed to improve the education of students at risk of educational failure because of limited English proficiency, poverty, race, geographic location, or economic disadvantage.

EXECUTIVE SUMMARY

This study reports the results of a four-year multi-method evaluation of the implementation of the Baltimore Curriculum Project (BCP) in six Baltimore City schools. BCP used a combination of the Direct Instruction (DI) program and Core Knowledge as its reform curriculum. Each of the six schools was demographically matched with a similar, within-district school so that it would have a reasonable control against which it could be compared. Two cohorts of students in the BCP and the control schools were followed through the course of the evaluation—students who were in either kindergarten or grade two during the 1996-97 school year (primarily in third and fifth grades, respectively, during 1999-2000). Interviews with principals and DI coordinators and focus groups with teachers were conducted each of the four years of the study to gauge BCP-school staff perceptions of the ongoing innovation. In the first three years of the study, detailed classroom-level observations were made in the BCP schools. Data collected provided evidence about the implementation and the classroom-level effects of the BCP curriculum.

Classroom observations and interviews indicated that the Direct Instruction curriculum and instructional methods were indeed implemented in the BCP schools, though the developer noted that implementation did not proceed at the desired rate in kindergarten until the fourth year. Implementation of Core Knowledge was not envisioned to begin until year 3, and proceeded more slowly than DI implementation. Teacher surveys and focus groups found positive views of both DI and Core Knowledge, but also revealed some frustrations.

Analyses of achievement test data indicated mixed results for students, depending on the subject, grade level, and school. In general, results were most positive for mathematics computation, though improvement in reading comprehension also occurred.

Mathematics computation scores rose dramatically at DI schools. Among the original kindergarten cohort, DI students moved, on average, from the 16th percentile at the end of first grade to the 48th percentile at the end of third grade (compared with growth among control counterparts from the 27th to 36th percentile over the same period). The impact on computation achievement for the original second grade cohort was nearly as strong. On the other hand, while DI students improved somewhat in mathematics concepts achievement, they continued to score well below national norms and their control counterparts in mathematics concepts (26th percentile).

Students at Direct Instruction schools also made considerable progress in reading over the course of the four years. On the primary measure of reading comprehension, members of the original kindergarten cohort were, on average, reading at grade level (49th percentile) by the end of third grade (after scoring, on average, at the 17th percentile on the readiness pretest, the Peabody Picture Vocabulary Test). Members of the original second grade cohort were nearing grade level (40th percentile) by the end of fifth grade. However, at the four schools with the highest rates of poverty and minority students, the average reading comprehension achievement was at the 38th percentile for the original kindergarten cohort and the 33rd percentile for the original second grade cohort. Students at control schools (where other curricula to improve reading achievement were being implemented) were achieving at the same level, so there were

no significant differences between the outcomes for the two groups (controlling for demographics and pretest factors).

Though limitations of the study make causal interpretations problematic, we view these findings as evidence that Direct Instruction (implemented at comparable levels of developer support) is a viable whole-school reform option for raising student achievement in reading and mathematics. While the reform may not necessarily perform better than other curricular alternatives, there have been sufficient achievement gains to justify its continuation as a reform option. In schools where teachers have become heavily invested in the program and scores are rising, we believe it is particularly important to continue implementing the reform, as change would be potentially disruptive. Based on the evidence from this four-year study, we would recommend that schools consider Direct Instruction as one of several reform options aimed at boosting student achievement, and make their choices based on the needs of their students and the capacities and preferences of their teaching staffs.

Annual evaluation reports were produced originally for The Abell Foundation. This final evaluation report is intended for the entire educational research community.

ACKNOWLEDGMENTS

The authors would like to thank the following people for their research assistance: Barbara McHugh, Florence Snively, Sam Kim, Tiffany Thornton, Laura Rachuba Overman, Gina Hewes, Marian Amoa, Judi Paver, Tiffany Meyers, Kristi Dusek, LuJu Lin, Kathleen Romig, and Naomi Graeff.

We also thank the following people who offered comments on earlier reports: Robert Slavin, Muriel Berkeley, Christopher Doherty, Gary Davis, Kurt Engelmann, Barak Rosenshine, and Gilbert Austin.

TABLE OF CONTENTS

Introduction 1

Design of Study 3

Results 5
- Evidence of Implementation 5
  - Integration of Direct Instruction and Core Knowledge 6
  - Perceptions of Baltimore Curriculum Project Faculty 8
  - BCP Curriculum 9
  - Direct Instruction Implementation 11

- Student Outcomes 13
  - Retention in Grade and Special Education Placement 13
  - Reading Achievement 14
  - Mathematics Achievement 20
  - Maryland School Performance Assessment Program (MSPAP) Outcomes 23

Discussion and Conclusion 27

References 29

Appendix 33

INTRODUCTION

Over the past decade, schools that serve students "placed at risk" because of poverty have been encouraged to implement programs with successful records of boosting student achievement. Funding from the New American Schools organization in the early 1990s helped to scale up implementation of several whole-school reform models and to evaluate these programs (Kearns & Anderson, 1996). Congressional approval of the Comprehensive School Reform Demonstration (CSRD) program in 1997 spurred more schools to implement whole-school reform models with proven track records of success, along with numerous studies of the impact of these programs (e.g., Borman, Hewes, Rachuba, & Brown, 2002). Though many studies have suggested that externally developed school reform models, or "promising programs," have advantages over locally developed reforms in systemically raising students' academic achievement (e.g., Herman et al., 1999; Nunnery, 1998; Stringfield et al., 1997), the nationwide whole-school reform movement has suffered setbacks recently as districts, such as Miami-Dade County, Florida, Memphis, Tennessee, and San Antonio, Texas, have abandoned their large-scale attempts to implement externally developed reform models after internal reviews (Viadero, 2001). A RAND study recently found significant gains for such whole-school reform models in just half the 163 schools under study (Berends, Kirby, Naftel, & McKelvey, 2001). The debate has continued as researchers have criticized the methodology used in internal district evaluations and the RAND study (Viadero, 2001). One of the salient issues in the debate has been the validity of school level data compared to longitudinal student level data. Another issue has been the need for more attention to the problems of implementation (Viadero, 2001). Rosenshine (2002) has also pointed out the need for more analysis of the various components that may be added to a whole-school reform model in a particular school setting.

In the context of the ongoing debate over the effectiveness of whole-school reform, additional longitudinal research studies that address both implementation issues and the effects of reform models on student achievement are particularly needed. More research on ways that various whole-school models are being combined in specific settings is especially important. The longitudinal study reported in this article contributes to that research base by analyzing the implementation and effects of a schoolwide reform model that joins two nationally disseminated and widely discussed models, Direct Instruction (DI) and Core Knowledge (CK), as the main components of the Baltimore Curriculum Project (BCP).

The Baltimore Curriculum Project

The Baltimore Curriculum Project was created and funded by educational reformers outside the Baltimore City Public School System (BCPSS) who believed that city students needed a more structured reading and writing curriculum. BCPSS student achievement on standardized tests during the early 1990s was substantially lower than in other parts of Maryland, with a majority of city students performing significantly below grade level in reading and mathematics (Lambert & Reynolds, 1997). Inspired by the successful implementation of the private Calvert Program in two Baltimore City public schools (McHugh & Spath, 1997; Stringfield, 1995), BCP sought to create a similar, structured curriculum from already existing curricular reform models. As the first director describes the process of creating BCP (Berkeley, 2002), the first programmatic

decision was to select the CK scope and sequence as a basis of what to teach (Core Knowledge Foundation, 1995; Hirsch, 1988; Hirsch, 1996). Though the Core Knowledge sequence specifically outlines a spiraling, detailed curriculum from kindergarten through eighth grade, it does not specify particular methods of instruction or an implementation strategy (Datnow, Borman, & Stringfield, 2000; Datnow, McHugh, Stringfield, & Hackler, 1998; Mac Iver, McHugh, & Stringfield, 2000). BCP staff felt that in order to deliver the curriculum, a more explicit instructional strategy was necessary to meet the needs of inner-city schools. After reviewing the available research on the effectiveness of various reform models for improving student achievement, BCP selected the DI model for reading, spelling, language arts, and mathematics. A body of research suggested that DI was an effective method of raising student achievement (e.g., Adams & Engelmann, 1996, and the studies discussed therein). BCP decided to create its own science and social studies units based on the Core Knowledge sequence, with specific lesson plans developed by BCP curriculum writers. The plan was to phase in CK lessons after implementation of DI was solidified and students were showing mastery of basic skills.

Implementation of BCP began in the fall of 1996 in six Baltimore City elementary schools. The school system agreed to support this externally initiated and externally funded reform effort, which expanded over the next several years and was institutionalized in the fall of 1998 as an alternative curriculum for 18 schools grouped into one administrative area (the "Direct Instruction Area"). The program was phased in gradually. During the first year (1996-97) of the program, DI reading and language arts were implemented in grades K-5 at four schools and grades K-2 at the two other schools (expanding to all grades in year 2). In addition, during the first year, a few teachers at each of the schools piloted CK lessons developed by BCP staff (though regular implementation of CK was not scheduled to begin until year 3). In year 2 (1997-98), DI spelling and mathematics were added to the implementation, though not at the higher grade levels at all of the schools. Pilot implementation of the CK curriculum continued, to various extents, in BCP schools during year 2. CK implementation expanded in years 3 and 4, but not at the level originally intended by BCP (an issue discussed more fully below).

While the DI program was designed for a full-day kindergarten program, it was not until year 3 that all of the original BCP schools had full-day kindergarten. Most of the six original BCP schools had a half-day kindergarten during year 1 of program implementation. Matched comparison schools were even less likely to have full-day kindergarten; three of the six control schools continued to have only half-day kindergarten during year 4 of implementation.

DI lessons are highly structured, teacher centered, and include careful sequencing and much repetition. Teachers use scripted lessons, which require no teacher development of lesson plans (although, to be effective, the lessons do require a certain amount of teacher preparation for lesson presentation). Much of the DI instruction takes place in homogeneous groups in which students are grouped according to current skill level. There is fluidity across groups, and students shift as necessary according to their performance. At BCP schools, instruction for DI reading and DI math was delivered by regular classroom certified teachers and, in some schools, classroom aides who received the same DI training as teachers.

To support the implementation of the BCP program, each school had a full-time BCP coordinator (often a "master-teacher" within the district) and designated teachers whose roles

were to be on-site grade-level "coaches" for the program. In addition, BCP arranged for each school to have a consulting relationship with the National Institute for Direct Instruction (NIFDI). Representatives from NIFDI provided a week-long training for DI reading and language arts instructors during the summer of 1996, before implementation began. Training for the DI program was extensive and ongoing, including school visits by NIFDI consultants each month. Training workshops for all components of the BCP program continued for teachers during each subsequent summer. The NIFDI representatives also reviewed paperwork and conferenced once a week with the DI coordinators in each school, and visited schools on a (generally) monthly basis to provide technical assistance for the DI portion of the BCP program. In the opinion of one prominent educational researcher (Rosenshine, 2002), this implementation of DI was atypical in its level of developer support to teachers and schools.

The relationship between NIFDI and one of the BCP schools ended in January 1998, when the school decided to drop the DI math curriculum (the original agreement was based on implementation of all DI programs). That school then selected another consulting firm that was amenable to the school's decision to drop the DI math curriculum and more flexible about making changes to the DI script. By June 1998, the relationship between another BCP school and NIFDI had ended as well, but that school also remained within the Direct Instruction administrative unit of the Baltimore City Public School System, and continued implementing DI under the supervision of that district office.

DESIGN OF THE STUDY

The Abell Foundation contracted with the Center for Social Organization of Schools (CSOS) of Johns Hopkins University for a multi-year evaluation of the implementation and outcomes of the Baltimore Curriculum Project in six Baltimore City schools. CSOS and Abell agreed that a multi-method study using data from achievement tests, observations, and interviews would be used to assess the effects of the new curriculum.

Sample of Schools

Each of the six schools was demographically matched with a similar, within-district school so that outcome comparisons could be made. (See Appendix for tables comparing BCP and control schools on demographic and other characteristics.)[1]

Sample of Students

Two cohorts in the BCP and control schools were followed through the course of the multi-year evaluation. These cohorts are students who were in either kindergarten or second grade during

[1] Because of differential demographic change in paired schools over the course of the evaluation study, as well as differences in demographic composition of particular paired cohorts, it is still necessary to control for differences between the BCP and control school cohorts in analyses.

the 1996-97 school year (primarily in third and fifth grades, respectively, during 1999-2000). Although it is possible to analyze outcomes for other cohorts receiving DI, these are the only cohorts for whom pretest or early covariate achievement measures are available from the first year of the study.

Process-Implementation Measures

Interviews with principals and DI coordinators and focus groups with teachers were conducted each of the four years of the study to gauge BCP-school staff perceptions of the ongoing intervention. In the first three years of the study, detailed classroom-level observations were made in the BCP schools.[2] Data collected provided evidence about the implementation and the classroom-level effects of the BCP curriculum. Classroom observations during year 4 of the study occurred when researchers shadowed DI coordinators (often with outside consultants as well) in each of the schools for a day.

Outcome Measures

The primary student achievement outcome measures used in this study were scores on the reading and mathematics subtests of the Comprehensive Test of Basic Skills, Fifth Edition (CTBS/5–TerraNova) (CTB/McGraw Hill, 1997). We used a curriculum-based measure (CBM), an individually-administered test of oral reading fluency, as a secondary outcome measure for reading. Covariate measures included the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1981), and the Comprehensive Test of Basic Skills, Fourth Edition (CTBS/4) (CTB, 1991).

The CBM reading inventories were individually-administered assessments of student oral reading fluency. These assessments were conducted in the spring of 1999 among second and fourth graders (as well as twice during the 1997-98 school year in all first and third grade classrooms) in the BCP and control schools. Passages read by students came from the DI Reading Mastery series and a popular elementary school anthology of literature.

The PPVT is a norm-referenced, picture identification test that is used nationally to obtain a measure of students' language ability. It is considered to be a good predictor of future success in reading and permitted us to control for any prior reading readiness "advantage" on the part of some of the students. The PPVT was administered in the 1996-97 school year to the cohort of kindergarten students at both the BCP and control schools.

The CTBS/4 is a norm-referenced, multiple-choice test that has been found in a variety of studies to possess reasonable psychometric properties. The two subtests of reading comprehension and mathematical concepts (the more nearly "higher order" subtests in the basic skills area) were administered to all second grade students in each BCP school in the fall of 1996. The second grade students in the BCP and control schools were tested with the CTBS/4 in

[2] The observation system was adapted from those in Schaffer & Nesselrodt (1993) and Stringfield et. al. (1997). It included measures of time on task and elements of good instruction identified in Stallings (1980) and Slavin (1987).

the spring of 1997. In subsequent years, the evaluation team used the results of the Baltimore City Public School System's (BCPSS) annual testing using the CTBS/4 (1998 and 1999) and the CTBS/4 or TerraNova (2000).

RESULTS

The study was designed to yield findings regarding both implementation issues surrounding the reform and its impact on student achievement. The following section summarizes implementation findings based on classroom observations, teacher surveys, and interviews with school-based personnel, as well as with staff associated with BCP. Then we present student achievement findings based on the Comprehensive Test of Basic Skills (CTBS), TerraNova, and the Maryland School Performance Assessment Program (MSPAP).

Evidence of Implementation

The Baltimore Curriculum Project brought together two reform models that on an intellectual, theoretical level are very different school improvement strategies. The research question of whether these reform models could be "successfully married" was complicated by that fact that BCP sought to use DI and CK as tools for reform, rather than to explicitly implement a combination of two externally developed reform models. As the following analysis of implementation issues indicates, it is clear that DI proved to be a more dominant tool as BCP sought to achieve its primary goal of basic skills instruction. Because it took longer to bring students to grade level on basic skills than BCP expected, the implementation of the CK social studies and science lessons developed by BCP was delayed, and implementation levels of CK were considerably lower than implementation of DI at first. The following sections present analyses of classroom observation data, perceptions of the constituents involved with program implementation, and our own analysis of the implementation of the BCP program.

During the first two years of BCP implementation, two- to three-day visits were conducted at each of the six experimental schools. Full-day observations were conducted in kindergarten and second grade classrooms during year 1, and in first and third grade classrooms during year 2. Data were gathered during classroom observations using a system built on the earlier work of Schaffer & Nesselrodt (1993) and Stringfield et al. (1997). It included a QAIT (Quality, Appropriateness, Incentive, Time) framework, developed by Stringfield and others (1997) and informed by the field work of Stallings (1980) and the effective schools research of Slavin (1987). Teacher practices specified by Direct Instruction (Engelmann & Madigan, 1996) as being vital to the DI program were also incorporated into the observation system.

The observation system in years 1 and 2 measured the occurrence of nine procedures required by the DI model at the beginning of each lesson, eleven behaviors central to the presentation and teaching techniques specified by the DI model, and five student behaviors expected in each DI lesson. Analyses of observation data indicate that most of these practices required by the DI model occurred in at least 80% of the lessons observed, even during the first

year of implementation (see Mac Iver, McHugh, & Stringfield, 1999 and Stringfield, McHugh, & Datnow, 1998 for a more detailed report of observation findings).

The purpose of the year 3 classroom observations shifted somewhat, and did not focus on the intricacies of DI. Instead, observations focused on the extent to which students were on task during instruction (time on task) and on the elements of effective teaching categorized and measured through the QAIT model.

In 1998-1999, one-hour observations were conducted in 24 DI reading classrooms, 20 DI math classrooms and 13 CK classrooms in second and fourth grades at five BCP schools. The school no longer with NIFDI did not allow continued observations.

Time on Task. During classroom observations, at eight-minute intervals, researchers recorded the number of students who were on task, off task, or waiting for instruction to continue. Not surprisingly, there was a higher on-task rate for reading instruction (which was implemented for a greater length of time over the course of the study) than for math and CK instruction. The average percentage of students on task in reading was 87% (ranging across schools from 84% to 91%). Rates of time on task averaged 80% for math (ranging from 59% to 90%) and 81% for CK (65% to 94%). Though direct comparisons with time on task rates from years 1 and 2 should be made with care (due to differences in types of classrooms observed and different observers in year 3 than the two previous years), it appears there was an increase in student time on task over time (see Mac Iver, McHugh, & Stringfield, 1999; Stringfield, McHugh, & Datnow, 1998).

QAIT. The QAIT model (Slavin 1987; Stallings, 1980) includes measures of ***Quality*** of instruction, ***Appropriateness*** of the difficulty of information to be learned, ***Incentives*** to learn, and the use of ***Time*** in the classroom. Every 16 minutes during each classroom observation, researchers recorded the presence or absence of QAIT components. An analysis of these observations reveals that teachers and students exhibited the desired behavior most often when the behavior was an explicit requirement of DI. Overall, classroom observations indicated a positive use of time and incentive structures in BCP classrooms as well as clear and often lively presentation of material. At the same time, there was less evidence of the quality of instruction that seeks to ensure deeper student understanding of concepts and more in-depth student engagement with material (see Mac Iver, Kemper, & Stringfield, 2000 for a more detailed discussion).

Integration of Direct Instruction and Core Knowledge

The original intent of the BCP developers was to phase in the CK component of the reform, after the DI component had solidified. The CK component was intended to focus primarily on social studies and science, while the rest of the curriculum areas would be covered under DI. Though CK social studies and science lessons (developed by curriculum writers on the BCP staff) were piloted during the first two years of the reform, they were scheduled for "regular" implementation in year 3 (after teachers' mastery of DI reading and language in year 1, and DI spelling and mathematics in year 2).

In general, we saw more evidence of CK implementation in schools where a greater proportion of students were reading at higher levels in the Reading Mastery curriculum and there was not a perceived need to add an additional DI reading period during time that otherwise would have been devoted to CK social studies or science lessons. During year 3, it was difficult for the research team to schedule observation of CK social studies and science lessons. At one school, CK was scheduled only 20 minutes per day in some classrooms and was not on the schedule at all in other classrooms. At several schools, periods that, on paper, were allocated to the CK curriculum were often used for test preparation or other non-CK topics. BCP staff concurred with the DI developer, NIFDI, that it was more important to schedule additional DI reading periods at some schools than to implement CK social studies and science instruction. Our year 4 discussions with NIFDI consultants indicated that, understandably, they had little interest in the CK component of the reform.

While BCP staff remained committed to incorporating a CK component as well as DI, the decision to add an extra DI reading period almost necessarily cut time from CK lessons. NIFDI representatives openly discussed with us the conflicts they have had with principals who wanted to implement CK instruction in social studies and science rather than add the second reading period. The school district administrator over the group of DI schools admitted that "Core is still being sacrificed at times to bolster the DI—but with that long-term goal of improving student reading and improving the math skills. But we hope that we can get to the point where, particularly in the upper grades, that we can really start to push the Core more uniformly" (Thrift, 2000). During year 4, the DI Area Office instituted periodic performance assessments based on Core Knowledge instructional units, which may have helped to increase the time spent in Core Knowledge lesson instruction in BCP classrooms. As we note below, however, teachers continued to report that time for Core Knowledge "is the first to be cut."

By contrast, we observed DI being implemented daily at all schools in all reading and math classrooms even beginning in year 1. Though we did not conduct formal observations during periods scheduled for language and spelling instruction, our informal observations and discussions with staff indicated daily implementation of DI in these subjects as well. Though the technical expertise of teachers in delivering DI varied considerably, the fact that observers saw DI implemented daily in all classrooms and subjects, with DI practices evident in more than 80% of lessons observed even during year 1, led us to characterize the level of DI implementation as relatively high.

One limitation of the study is our lack of access to systematic data about additional interventions received by students at both program schools and comparison schools. Students may have received instruction in after-school programs, summer school, or other interventions during the day. For example, one of the program schools had an after-school program serving more than 100 children, which included a "100-book challenge" and book-club discussion format as well as other enrichment activities. Other schools may have had similar interventions in addition to the DI program.

Four of the six original BCP schools continued their association with the National Institute for Direct Instruction (NIFDI), the organization led by the original developer of the

Direct Instruction model.[3] The two schools that were not willing to implement all dimensions of the whole-school reform program, as defined by NIFDI, did continue implementation of the Reading Mastery curriculum, though not according to the exact specifications of the original developer. The one school to which we continued to have access for observation was given high marks in implementation, according to its consultant (JP Associates), and our observations concurred.

By contrast, the DI reform model developer team did not consider implementation levels to be high until year 4. And they viewed implementation at the fourth of NIFDI's remaining schools to be endangered, primarily because the current principal at that school (the third since implementation began) was not committed to following all of NIFDI's recommendations and staff turnover was very high (Davis, 1999). The key implementation problem in Baltimore, according to NIFDI, was the kindergarten program. While the Direct Instruction program was designed for a full-day kindergarten, it was not until year 3 that all the original BCP schools had a full-day kindergarten[4] (most of the six original BCP schools had a half-day kindergarten during year 1 of the program implementation). There was also a high turnover rate of kindergarten teachers, partially due to opposition to the program. By year 4 of the program, however, a NIFDI representative voiced optimism about all four of the BCP kindergarten programs (though they did not view the fourth as "highly implementing"). In addition, the developer judged DI implementation as low at schools where principals balked at scheduling extra DI reading periods during time scheduled for CK social studies and science lessons.

Because the technical requirements of full DI implementation, as defined by the developer, appear to be so heavily dependent on coverage of a certain number of lessons during kindergarten and the willingness of schools to implement "double reading periods" for students in subsequent grades, this four-year evaluation can only assess the early impact of reform before implementation problems had been fully worked out. It is important to note that the reform is likely to face similar "start-up" implementation problems in other urban settings, and that the technical requirements of implementation are an important factor for decision makers to consider. In addition, if high implementation levels of DI sometimes require all hours of the school day, it becomes difficult to successfully combine DI with another reform.

Perceptions of Baltimore Curriculum Project Faculty

Beyond conducting low- and high-inference classroom observations, members of the research team conducted focus groups with teachers and interviews with DI coordinators and principals at the BCP schools. In general, we followed the study cohorts over time, conducting focus groups with their teachers each year (kindergarten and second grade teachers during year 1, first and third grade teachers during year 2, second and fourth grade teachers during year 3, and third and fifth grade teachers during year 4). In addition, we obtained access to results of the teacher

[3] Given this issue of how fully Direct Instruction was implemented, analyses of student achievement (reported in upcoming sections) were conducted for the group of four NIFDI schools, as well as the full group of six.

[4] Three of the six matched control schools continued to have half-day kindergarten even during year 4 of the implementation.

survey conducted by the Baltimore Curriculum Project in January 1999. Results were available from teacher surveys at four of the six original BCP schools.

Focus group interviews are considered to be a highly efficient method of gathering data from a large number of people at once. They also give rise synergistically to insights and solutions that might not otherwise arise in single-person interviews (Brown, Collins, Duguid, 1989). In addition, focus group participants tend to provide checks and balances for each other, which in turn reduces problems with false or extreme views. The group dynamics of the situation also allow the researcher to determine if there are fairly consistent views among the participants (Patton, 1990).

Transcripts of the school interviews were analyzed using the constant comparative technique (Lincoln & Guba, 1985). The transcribed interviews were first unitized by segmenting the data into discrete, heuristic units. The individual units were then analyzed in terms of the themes that emerged from the data and placed into distinct and internally consistent categories.

The following is a summary of findings from the interviews and teacher surveys, organized around the four main themes that emerged from the data: BCP curriculum, DI implementation, relationship with DI consultants, and professional development. The comments have been aggregated across school and respondent level to ensure the confidentiality of all respondents. The group of survey respondents and the group of teachers participating in focus groups are not identical groups. While there is some overlap, not all survey respondents participated in focus groups, and not all focus group participants returned surveys. The summary reflects the wide diversity of opinions expressed by teachers.

BCP Curriculum

Analysis of the four years of focus group data indicated widespread teacher support for the systematic nature of the program. Because the BCP curriculum incorporates two very different components (DI and CK), and attitudes also varied depending on the particular program within DI, we analyzed responses from school-based staff with regard to (1) DI reading, (2) DI language and spelling, (3) DI math, (4) more general views regarding DI, and (5) CK.

Reading. Teachers viewed the DI reading program as the most effective component of the BCP curriculum, with two-thirds (68%) indicating on the 1999 survey that they found the DI reading lessons "very effective" and 30% responding "somewhat effective." (The BCP-designed survey used a three-point scale: very effective, somewhat effective, not effective.) Most of the focus group participants felt that the DI reading program gave children a firm foundation in reading skills. A number of teachers voiced their concern, however, that DI was not as appropriate a program for older elementary students, especially in comprehension.

DI Language and Spelling. Teachers were also generally positive about the language program. The majority of survey respondents (54%) said DI language lessons are "somewhat effective;" most of the rest (31%) viewed this program as "very effective." Though just a few (7%) judged it as "not effective," teachers did raise some concerns. Several teachers commented during focus groups that they do not believe the DI language program is the best way to teach

writing or to prepare students for writing on the MSPAP (which assesses various writing genres, such as writing to persuade). Upper-grade teachers would have preferred more integration between the DI reading and language programs so that students could respond in writing to literature. Comments about the spelling program were generally positive, though some teachers voiced concerns about too little attention to the meaning of words (which is covered in the DI reading program rather than the spelling program).

Math. Teachers were generally positive about the math program, with some reservations. One in four survey respondents (23%) viewed the math program as "very effective," and most of the rest (63%) saw it as at least "somewhat effective." While some teachers and administrators praised the math program (especially in its approach to word problems), many teachers did not believe the auditory nature of the program worked well for all children. Focus group participants at all of the schools emphasized the need to add manipulatives and other techniques to help students grasp mathematical concepts, and some teachers commented that the wording of the math script was hard for children to understand and needed adaptation. In addition, there was concern that the program-specific language and techniques used in the DI math program were not generalizable to non-DI situations (including standardized tests), and that some concepts were not presented early enough to prepare students for the CTBS and MSPAP. Several of the DI coordinators appeared to feel the math program was still in need of developer revisions (currently in process). The fix-ups for the math program are not built into the program as they are for reading, thus teachers have to be very secure with the lessons "to understand how to implement the corrections."

Overall Views about Direct Instruction. Across the four schools responding to the teacher survey, the large majority of teachers (75%) said that student achievement had increased since their school began using DI (ranging from 46% at one school to 100% at another). Roughly three in four teachers (77%) voiced support for continuing to use DI at their school (about one in four "enthusiastically" supported it).

All of the interviewed principals and DI coordinators expressed support for the DI program, and commented that they find the main strengths of the DI curriculum to be the structure and continuity of the program, as well as "the built-in ability for teachers to be constantly assessing students." Principals and DI coordinators felt the structure of the program offers both teachers and students a logical progression through reading, language, and math curricula, each of which builds sequentially upon itself. In addition, they viewed the positive behavior system with established rewards and consequences as another beneficial aspect of the entire program.

Though teachers saw positive results and thought the program was working, there were many concerns voiced during focus group interviews. Frustration about the "robotic" nature of the program was mentioned during focus groups at more than one school, and teachers voiced some doubts about the ability "of kids to transfer the DI knowledge to other situations." Though many administrators did not like being driven by MSPAP, teachers repeatedly maintained that the DI curriculum (especially language and math) does not teach "enough MSPAP skills." This appeared to reflect an underlying concern among teachers that DI does not address conceptual thinking skills as well as it addresses more basic skills.

Core Knowledge. A number of teachers and administrators consciously perceived the CK component of the BCP curriculum as complementary to DI and designed to address some of the MSPAP skills. Most teachers felt the CK curriculum was a useful tool for engaging students and building their general knowledge base, but many had difficulty actually implementing the CK lessons. Results from the teacher surveys indicated that about half of the responding teachers used CK lessons two times or less a week. One of the original BCP schools allotted just 20 minutes a day for CK in some classrooms, and none in other classrooms.

Most teachers reported that finding time to work in the CK lessons was difficult, and that "Core is the first to get cut when time is tight." Schools that had implemented double DI reading remarked that there was even less time for CK. Teachers complained that too much material was compressed in the CK lessons from BCP, making it difficult to actually cover all of the activities in one period. Only 23% of the teachers who responded to the survey reported that they used all of the CK lessons for their grade. Teachers also did not perceive DI and CK to be well integrated. As one third grade teacher put it during a focus group: "There is no marriage between Core and DI. They are separate and distinct. So there is DI and then there is Core and there is no connection."

Direct Instruction Implementation

With few exceptions, all the interview respondents (principals, DI coordinators, and teachers) perceived the overall implementation of DI at their school to be well underway by the third year, with teachers more skilled in DI techniques, and instruction improving. At the same time, administrators commented that faculty turnover and student transience made maintaining a stable DI program difficult. Principals stated that some of the turnover was due to teachers who disagreed with the program and preferred to go elsewhere. By the fourth year of implementation, however, the mobility of teachers at BCP schools had decreased, compared to previous years (Thrift, 2000).

Respondents generally felt pressure to cover the curriculum materials as quickly as possible. This stress on lesson coverage appeared to reflect an "urgency to get kids' skills up," as one respondent put it. While the coordinators and principals stressed the need for coverage with mastery, teachers expressed fear of being "placed back" and forced to re-teach lessons, which they said was often frustrating and demoralizing for their students as well as for themselves.

Many teachers expressed frustration at the lack of flexibility of the DI program and the "boring" scripts for the lessons. While coordinators emphasized that there were acceptable ways to deviate and expand upon DI lessons, teachers appeared neither sure when such deviations were appropriate nor confident about what constituted "acceptable deviations." For the first few years, they said they were told to "just stick to the script" and then, in the fourth year, they were told that deviations to enhance student understanding were necessary. The overwhelming theme of coordinator responses to their frustration was that "teachers must first learn to be good technicians [master the DI process] before they can become engineers [make deviations from the DI script]." Teachers also confessed that they felt "guilty" when they reverted to "traditional teaching methods" to get a concept across to students, but felt that methods other than DI were sometimes necessary to facilitate student learning.

Relationship with DI Consultants. As described earlier, consultants from the National Institute for Direct Instruction (NIFDI) visited classrooms in each school each month to evaluate how teachers were implementing DI and provide modeling and feedback on their teaching. The relationship between teachers and the consultants, which was the source of considerable tension in earlier years of BCP implementation, appeared to improve over time. A universal theme from the first three years of teacher focus group data was that the consultants did not treat the teachers as professional educators and tended to discount their years of experience in the classroom. The majority of teachers felt mid-lesson corrections (in which the consultant would interrupt the class and take over teaching when DI techniques were not being executed properly), were "unprofessional" and "unnecessary." In addition, most teachers said that the feedback they received from consultants tended to be mostly negative. Several objected to an "outsider" making supervisory and instructional decisions at their school. By the fourth year, many of the original consultants had been replaced, and there was a general consensus that the relationship with the consultants had improved.

Professional Development. Professional development sessions occurred much more frequently during the first two years of the BCP implementation than in the later years, when funding was reduced and there was less time allocated by the district. Veteran teachers mentioned that by the third year of implementation the training had become repetitious. By contrast, some of the newer faculty members felt they would have benefited from more coaching. While teachers at all schools mentioned that they had teachers who serve as DI coaches, almost all of them said that there were few opportunities for them to either observe the coaches or to be observed by the coaches. Teachers at several schools also commented that they felt as if they could use more intensive instructional support for delivering the CK component of the BCP curriculum. Teachers at all schools commented that they felt that they were "missing out" on workshops and in-services offered to teachers at non-DI schools, and, as a result, were missing valuable information and technical skills that could benefit them as teachers, not just as DI teachers.

Relationship between BCP Schools and the District Office. During the first two years of DI implementation, several of the schools were in the same geographic district under an area executive officer (AEO) who was eventually asked to head the new Direct Instruction Area. Principals from these schools, as well as others in different administrative areas, generally reported receiving support from their supervisors in their implementation of Direct Instruction, though the degree of *active* support varied. Sometimes principals viewed "support" as simply leaving the school alone. Schools were generally able to receive exemptions or waivers from district policies (such as quarterly assessment requirements) without too much hassle, and one AEO (other than the one who eventually headed the DI area) actually intervened to stop a principal who was administering assessments from which the school could have been exempted.

Though they were generally able to implement DI without interference, principals perceived a general lack of support from most central office administrators. As one principal said in the spring of 1998, "The curriculum officers are not [supportive].... It's like a stepchild. And I can understand where they're coming from. They have curriculum that's purely constructivist, and this is the total opposite end of the spectrum." Others noted that no one from the central office (other than the AEO) had come to observe the DI reform in action.

The creation of a DI administrative unit at the district office had pros and cons for administrators at the original six BCP schools. Principals appeared to feel that the personnel at the DI area office were quite helpful. They perceived fewer conflicts with the central office over district-wide policies related to such issues as professional development and assessments. But interaction with the DI area office was somewhat frustrating for the coordinators at the BCP schools. As all of the BCP schools were using the DI program before the area office was created, staff members at these schools perceived themselves to be at the high end of the learning curve with regard to program implementation. Thus, some of the mandatory paperwork and in-services "did more to hinder than to help" the BCP schools. Teachers' perceptions of the assistance provided by the DI area office in years 3 and 4 of implementation varied greatly from school to school, with teachers at one school stating that "DI area people are here all the time," while teachers at another school were not sure exactly who the DI area office personnel were.

Student Outcomes

In the following sections, we examine the effects of this reform effort on several student outcomes. Measures include: retention in grade, special education placement, student-level achievement in reading and mathematics (measured by CTBS scores), oral reading fluency, and school-level achievement on the Maryland School Performance Assessment Program (MSPAP).

Retention in Grade and Special Education Placement

Analyses of student outcomes after four years of Direct Instruction implementation showed that students in these schools were less likely to be retained in grade (or more likely to be promoted) than their control counterparts. In the original second grade control cohort, just 1% of those remaining at the same BCP school after four years were in fourth rather than fifth grade in the spring of 2000, compared with 16% of the control cohort retained. Among the original kindergarten cohort, 21% of the control students were in second rather than third grade in the spring of 2000, compared with only 4% of the BCP cohort retained. Due to limitations in the available data, we are unable to calculate trends in retention rates over a four-year period at study schools prior to the introduction of the reform. It is possible that the schools that adopted Direct Instruction had lower retention rates than their comparison schools before introducing the reform, so we cannot conclude with certainty that Direct Instruction is responsible for a lower retention rate. It is likely, however, that the reform helps to account for this differential retention rate, because the structural characteristics of Direct Instruction allow for regrouping of students for reading instruction so that students may receive instruction at a lower grade level without being formally retained.

Differences between BCP and control schools in assignment of students to special education were not as pronounced. Among the original kindergarten cohort, 5% of BCP students were assigned to special education in 1999-2000, compared with 6% in control schools. Among

the original second grade cohort, 14% of BCP students had special education status, compared to 10% in control schools.[5]

Reading Achievement

The first question we address in this section is the evidence of growth in reading achievement for the study cohorts over the four years of this study. Though the size of the original cohorts declined by at least half over the four years, we first consider those students who received four years of Direct Instruction in reading, compared with control students who remained at the same school for the four years of the study. Sample mortality does not appear biased except for retention rates (which we adjust for by including retained students in the scale score analysis). There were no significant differences between DI and control schools in the readiness or achievement levels of students who were lost due to mobility. Students appear to be transferring for reasons unrelated to academic achievement.

The control schools used a variety of basal readers during the first two years of this study. The extent of phonetic instruction varied from school to school and from classroom to classroom. In the fall of 1998 (the third year of this study), the school system shifted to the Open Court reading series (another highly prescriptive, highly phonetically-based early reading program) for grades K-2. Control members of the original kindergarten cohort thus received instruction using Open Court during their second-grade year[6] at five of the six control schools.[7] Beginning in the fall of 1998, the school system adopted the Houghton Mifflin reading series for grades 3-5, and control students in the original second grade cohort (who were generally fifth graders in 1999-2000) had neither a strong phonetically-based reading program in the earlier grades, nor a reading program that necessarily provided opportunities for them to acquire phonetic training or word attack skills in later grades.

Tables 1 and 2 summarize NCE gains in reading for the original cohorts.[8] Even though the tests are not strictly comparable, these gains give a reasonable estimate of how much reading growth occurred for each group over the four-year period. For the original kindergarten cohort, we report the average NCE scores on the Peabody Picture Vocabulary Test, the first grade CTBS/4 reading tests, and the third grade CTBS/5 reading tests.[9] For the original second grade cohort, we report the average NCE scores on the second grade CTBS/4 reading comprehension test (Spring 1997), third grade CTBS/4 reading tests (Spring 1998), and the fifth grade CTBS/5 reading tests (Spring 2000).[10]

---

[5] Given the changes in how special education data were reported over the time period of the study, we do not attempt to present trend tables for special education.

[6] Retained control students received both first and second grade instruction using Open Court.

[7] Control School 3 was part of a small group of schools in the district that continued to use Houghton-Mifflin in the primary grades.

[8] Scale score conversions for the CTBS/4 to CTBS/5 were not available when these analyses were conducted.

[9] Retained students, who were in second rather than third grade in the spring of 2000, were analyzed separately since NCE scores correspond to particular grade level versions of the test.

[10] Retained students, who were in fourth rather than fifth grade in the spring of 2000, were analyzed separately since NCE scores correspond to particular grade level versions of the test. Spring 1999 scores are not reported since they are not available for all schools in the study.

Table 1. Mean NCE Reading Scores for Original Kindergarten Cohort at BCP and Control Schools, Fall 1996–Spring 2000 [11]

School (Number of students)	*1996-97 Peabody Picture Voc. Test NCE* (Percentile)	*Spring 1998 CTBS/4 Reading Vocabulary NCE* (Percentile)	*Spring 1998 CTBS/4 Reading Comp. NCE* (Percentile)	*Spring 2000 CTBS/5 Reading Vocabulary NCE* (Percentile)	*Spring 2000 CTBS/5 Reading Comp. NCE* (Percentile)
All BCP Schools (n=171)[12]	29.7 (17th)	41.3 (34th)	40.7 (33rd)	46.5 (43rd)	49.3 (49th)
All Control Schools (n=104)	31.6 (19th)	47.8 (46th)	42.9 (37th)	51.9 (53rd)	51.6 (53rd)

Table 2. Mean NCE Reading Scores for Original Second Grade Cohort at BCP and Control Schools, Spring 1997–Spring 2000

School (Number of students)	*Spring 1997 CTBS/4 Reading Comp. NCE* (Percentile)	*Spring 1998 CTBS/4 Reading Vocabulary NCE* (Percentile)	*Spring 1998 CTBS/4 Reading Comp. NCE* (Percentile)	*Spring 2000 CTBS/5 Reading Vocabulary NCE* (Percentile)	*Spring 2000 CTBS/5 Reading Comp. NCE* (Percentile)
All BCP Schools (n=182)	39.1 (30th)	39.5 (31st)	42.9 (37th)	48.5 (47th)	44.9 (40th)
All Control Schools (n=132)	38.1 (29th)	43.8 (38th)	43.2 (37th)	45.7 (42nd)	45.7 (42nd)

Overall, the students in the kindergarten study cohort began school below average in reading readiness, as measured by the Peabody Picture Vocabulary Test. By the end of third grade they were, on average (including the school with low FRL rates), reading at about grade level (49th percentile). Examining the data from a slightly different perspective, we find that nearly half of the remaining members of the original kindergarten cohort at both BCP schools (46%) and control schools (45%) were reading at the third grade level (50th percentile) or above by the spring of 2000.[13] These results are influenced, however, by large numbers of students at the relatively advantaged BCP school. When just the four NIFDI schools and their controls are included in the analysis (excluding the school with the lowest free lunch rate), the average reading comprehension is at the 38th percentile, and the average reading vocabulary is at the 33rd percentile. At these high-poverty schools, we find 35% of the original kindergarten BCP cohort reading at grade level or above by third grade, compared to 40% in the control cohort.

[11] Spring 1999 scores are not available for all schools, and so this column is omitted.

[12] Fewer students in both cohorts also took the vocabulary subtests.

[13] This analysis is based on the scale score associated with 50th percentile for third grade, and includes retained second-graders as well as the third graders. Though there was little difference overall, further analyses suggest that DI schools had a somewhat greater percentage of boys reading at grade level than control schools did (and a somewhat lower percentage of girls reading at grade level than at control schools). Unfortunately, small sample sizes make it difficult to explore this interactive effect of program x gender.

2 4

Students who received four years of Direct Instruction beginning in second grade were nearing grade level (approximately the 40th percentile) by the end of fifth grade (with students at the four high-poverty schools at the 33rd percentile). Even when retained students are included in analyses, there was virtually no difference in reading achievement between the groups. Among the students who remained in the same school for the four-year period, 39% of BCP students (compared with 38% of the control cohort) were reading at fifth grade level or above in the spring of 2000.

To address the question of whether Direct Instruction produced significantly better achievement outcomes than instruction at control schools (controlling for pretest measures and demographic variables), we conducted regression analyses of spring 2000 reading comprehension and reading vocabulary scale scores[14] for the original 1996-97 kindergarten and second grade cohorts (primarily in third and fifth grades at the time of testing). The analyses include retained students in second and fourth grades, respectively, a notably larger group in control schools than BCP schools (as discussed previously). Only students who remained at the same school (Direct Instruction or control school) for the four years were included. We conducted analyses with all six original pairs of schools, as well as with the four pairs of schools that continued with the original consulting group (NIFDI). (Though all schools continued to implement DI reading, implementation at two of the original schools differed slightly after they changed consultants.)

For the original kindergarten cohort, we examined the effect of four years of Direct Instruction on both reading comprehension and reading vocabulary scores, controlling for pretest scores on the Peabody Picture Vocabulary Test and demographic variables. Over that period, there was a marginal effect of Direct Instruction on reading comprehension scores ($p=.14$, effect size=.13 for six pairs; $p=.13$, effect size=.16 for four pairs).[15] There was no measurable effect on reading vocabulary scores. In other words, the reading achievement of Direct Instruction students was neither significantly better nor worse than that of control students.

Because pure pretest scores for the original second-grade control cohort are not available, our analysis of effects on that cohort controls for spring 1997 reading comprehension test scores (after the first year of Direct Instruction implementation) as well as demographic variables. The analysis includes retained students, but only those students who have remained at the same school over the four years. Over the three-year period (after the first year of Direct Instruction), there was a non-significant effect on reading vocabulary scores ($p=.166$, effect size=.14 for six pairs; $p=.10$, effect size=.21 for four pairs). There was no significant effect of Direct Instruction on reading comprehension scores for the three-year period (though there could have been an effect during the first year of Direct Instruction that we are not able to detect with the data available).

---

[14] The original design of the study included only measures of reading comprehension, and in the spring of 1997, when the Baltimore City Public School System did not administer the CTBS to most elementary students, the Johns Hopkins University research team administered only the reading comprehension subtest to the second grade cohort. Since later reading vocabulary scores are available, we include them in analyses as another measure of reading. We maintain, however, that the reading comprehension measure is the most important (Stanovich, 1991; Daneman, 1991).

[15] These results meet neither the usual standard for statistical significance ($p < .05$) nor the usual standard for a meaningful effect size (.25, or one-quarter of a standard deviation).

Because sample sizes were so greatly reduced over the study, we also analyzed test score results for the full group of third and fifth graders at DI and control schools in 1999-2000, many of whom had not been at the school for the full four-year period. We conducted regression analyses, controlling for previous year's reading score, demographic variables (race, sex, free-lunch status), attendance, and mobility (whether at the same school as the previous year). Though such an analysis is able to detect only a one-year effect of Direct Instruction on reading achievement, it provides useful information for urban school districts with particularly mobile student populations. The effect of one year of Direct Instruction on reading vocabulary scores was significant at the fifth grade level (p=.002, effect size=.24). Direct Instruction did not, however, have a positive one-year effect on reading comprehension scores at the fifth grade level, and had no one-year effect on either vocabulary or comprehension scores at the third grade level.

Tables 3 and 4[16] below show the sizes of the NCE gains over the year (not controlled for demographic differences and pre-existing achievement differences). Taken together with the multivariate regression analysis that controls for these differences, the results suggest that both DI and control schools were making positive progress in raising student reading achievement. Average achievement moved closer to grade level (50th percentile) at both sets of schools, though it did not reach grade level in this study.

Table 3. Mean CTBS Scores and Gain Scores in Reading Comprehension and Vocabulary for Third Grade Cohort at BCP and Control Schools, Spring 1999–Spring 2000
(Not Necessarily Same School in Spring 1999 and Spring 2000)

	Spring 1999 (2nd Grade)		*Spring 2000* (3rd Grade)		*Comp. Gain*	*Voc. Gain*
School (Number of students)	*Comp. Mean NCE* (Standard Dev.)	*Voc. Mean NCE* (Standard Dev.)	*Comp. Mean NCE* (Standard Dev.)	*Voc. Mean NCE* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)
All BCP Schools (n=240)[17]	40.3 (18.7)	42.6 (22.3)	47.6 (21.7)	45.5 (20.9)	7.3 (14.5)	2.9 (15.3)
All Control Schools (n=240)	38.9 (16.7)	41.5 (21.0)	45.2 (19.9)	44.7 (20.2)	6.3 (14.5)	3.2 (16.3)

[16] These analyses include five of the six pairs of schools. One pair is excluded because complete data from the spring of 1999 are not available.
[17] Somewhat fewer students took both vocabulary tests.

Table 4. Mean CTBS Scores and Gain Scores in Reading Comprehension and Reading Vocabulary for Fifth Grade Cohort at BCP and Control Schools, Spring 1999–Spring 2000
(Not Necessarily at Same School in Spring 1999 and Spring 2000)

School (Number of students)	*Spring 1999* (4th Grade)		*Spring 2000* (5th Grade)		*Comp. Gain*	*Voc. Gain*
	Comp. Mean NCE (Standard Dev.)	*Voc. Mean NCE* (Standard Dev.)	*Comp. Mean NCE* (Standard Dev.)	*Voc. Mean NCE* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)
All BCP Schools (n=239)	45.8 (18.5)	40.6 (22.2)	44.9 (20.1)	47.5 (20.5)	-0.9 (11.5)	6.9 (15.8)
All Control Schools (n=199)	39.7 (17.8)	36.8 (20.6)	41.2 (19.3)	40.2 (18.1)	1.5 (14.3)	3.4 (17.9)

The Effect of DI Reading on Mobile Students. Since mobility in urban districts is a pressing issue that affects achievement (Kerbow, 1996) and previous research has suggested a particularly positive impact of DI on mobile students (Brent & DiObilda, 1993), we also sought to determine whether Direct Instruction had a particularly useful impact on mobile students in this study. In a more preliminary report on this study (Mac Iver, Kemper, & Stringfield, 2000), we found a significant effect of DI instruction on one-year gains in reading comprehension for fourth graders new to a study school (mobile transfer students). Those fourth graders new to DI schools (n=29) gained an average of 6.4 NCE points in one year (from 31.8 to 38.2), compared with a gain of just 0.4 NCE points (from 35.4 to 35.8) for control students (n=46).

Seeking to determine whether such an effect on new students could be replicated, we examined students new to the study schools in 1999-2000 (who had transferred from another city school and had a test score from the previous year). Tables 5 and 6 present the NCE gains for these students who had one year of Direct Instruction in third and fifth grades, respectively. New students at DI schools did have higher gains than students at control schools, but these differences were not significant.[18] It is also important to note that new fifth grade students came into DI schools with significantly higher reading achievement scores than the new students in control schools, whereas new third graders in DI schools were slightly below new third graders in control schools.

[18] This could be due to the relatively small sizes of these groups of new students.

Table 5. Mean CTBS Scores and Gain Scores in Reading Comprehension and Reading Vocabulary for New Fifth Grade Cohort at BCP and Control Schools, Spring 1999–Spring 2000
(Not at Same School in Spring 1999 and Spring 2000)

	Spring 1999 (4th Grade)		*Spring 2000* (5th Grade)		*Comp. Gain*	*Voc. Gain*
School (Number of students)	***Comp. Mean NCE*** (Standard Dev.)	***Voc. Mean NCE*** (Standard Dev.)	***Comp. Mean NCE*** (Standard Dev.)	***Voc. Mean NCE*** (Standard Dev.)	***Mean NCE Gain*** (Standard Dev.)	***Mean NCE Gain*** (Standard Dev.)
All BCP Schools (n=22)	38.9 (18.9)	37.0 (21.2)	41.9 (19.4)	42.6 (21.7)	3.0 (9.0)	5.6 (15.5)
All Control Schools (n=23)	29.3 (14.0)	22.0 (20.6)	29.7 (19.3)	28.8 (12.7)	.5 (12.1)	6.8 (12.1)

Table 6. Mean CTBS Scores and Gain Scores in Reading Comprehension and Reading Vocabulary for New Third Grade Cohort at BCP and Control Schools, Spring 1999 – Spring 2000
(Not at Same School in Spring 1999 and Spring 2000)

	Spring 1999 (2nd Grade)		*Spring 2000* (3rd Grade)		*Comp. Gain*	*Voc. Gain*
School (Number of students)	***Comp. Mean NCE*** (Standard Dev.)	***Voc. Mean NCE*** (Standard Dev.)	***Comp. Mean NCE*** (Standard Dev.)	***Voc. Mean NCE*** (Standard Dev.)	***Mean NCE Gain*** (Standard Dev.)	***Mean NCE Gain*** (Standard Dev.)
All BCP Schools (n=19)	32.4 (20.3)	35.0 (22.6)	41.4 (19.6)	42.5 (22.7)	9.0 (18.1)	7.5 (14.4)
All Control Schools (n=40)	35.9 (16.2)	36.7 (19.2)	41.7 (18.2)	41.5 (20.0)	5.8 (12.0)	4.8 (18.4)

Oral Reading Fluency. As some have argued for considering the results of more curriculum-based testing or individualized testing of student reading ability (Deno, 1985; Hall & Tindal, 1989; Fuchs & Deno, 1992; Hasbrouck & Tindal, 1991; Marston & Deno, 1982; Marston, Deno, & Tindal, 1983), we also conducted individual tests of oral reading fluency among study cohorts during the second and third years of the study. Regression results using spring 1999 Individualized Reading Inventory test scores as the dependent measure of student achievement (controlling for 1998 CTBS/4 reading comprehension score, demographic variables, 1998-99 attendance, and whether student was at the same school as last year) indicate that Direct Instruction had a significantly positive effect on oral reading fluency at both the second grade ($p=.024$, effect size=.15) and fourth grade ($p<.0005$, effect size=.26) levels. When analyses are restricted to students who have been in the same school for three years, effect sizes remain similar, but the effect was no longer significant at the second grade level (Mac Iver, Stringfield, & Hall, 1999).

Mathematics Achievement

The first question we address in this section is the evidence of growth in mathematics achievement for the study cohorts over the study. We first consider those students in the original cohorts who received three years of Direct Instruction in mathematics (beginning in the fall of 1997),[19] compared with control students who remained at the same school for the four years of the study. The schools included in these analyses are also those with the highest poverty levels.

Tables 7 and 8 summarize NCE gains in mathematics for the original cohorts. Even though the tests are not strictly comparable, they give a reasonable estimate of how much mathematics growth occurred for each group over the three-year period. For the original kindergarten cohort, we report the average NCE scores on the first and second grade CTBS/4 math tests, and the third grade CTBS/5 math tests.[20] For the original second grade cohort, we report the average NCE scores on the second grade CTBS/4 mathematics concepts test (Spring 1997), third and fourth grade CTBS/4 math tests (Spring 1998 and 1999), and the fifth grade CTBS/5 math tests (Spring 2000).[21]

Table 7. Mean Math NCE Scores for Original Kindergarten Cohort at BCP and Control Schools Spring 1998–Spring 2000 [22]

School (Number of students)	*Spring 1998 CTBS/4 Math Concepts NCE* (Percentile)	*Spring 1998 CTBS/4 Math Computation NCE* (Percentile)	*Spring 1999 CTBS/4 Math Concepts NCE* (Percentile)	*Spring 1999 CTBS/4 Math Computation NCE* (Percentile)	*Spring 2000 CTBS/5 Math Concepts NCE* (Percentile)	*Spring 2000 CTBS/5 Math Computation NCE* (Percentile)
All BCP Schools (n=104)[23]	33.0 (21st)	28.0 (16th)	35.1 (24th)	40.9 (33rd)	36.4 (26th)	48.8 (48th)
All Control Schools (n=67)	44.3 (39th)	37.2 (27th)	38.7 (30th)	40.5 (33rd)	43.8 (38th)	42.7 (36th)

[19] Since Schools 1 and 5 did not have uninterrupted implementation of DI Mathematics since the fall of 1997, they and their control schools are excluded from the following analyses. In addition, since one of the remaining four BCP schools (School 2) implemented DI Mathematics only in grades K-2 during 1997-98 (so that the original second grade cohort at that school did not receive treatment that year), that school and its control are not included in the analyses for the original second grade cohort.

[20] Retained students, who were in second rather than third grade in the spring of 2000, were analyzed separately as NCE scores correspond to particular grade level versions of the test. This analysis is based on the four schools that implemented DI Math for three years for this cohort of students.

[21] Retained students, who were in fourth rather than fifth grade in the spring of 2000, were analyzed separately as NCE scores correspond to particular grade level versions of the test. Spring 1999 scores are not reported because they are not available for all schools in the study. This analysis is based on the three pairs of schools that implemented DI Math for three years for this cohort of students.

[22] Spring 1999 scores are not available for all schools, and so this column is omitted.

[23] Fewer students also took the vocabulary subtests.

Table 8. Mean Math NCE Scores
for Original Second Grade Cohort at BCP and Control Schools
Spring 1997–Spring 2000

School (Number of students)	*Spring 1997 CTBS/4 Math Concepts NCE* (Percentile)	*Spring 1998 CTBS/4 Math Concepts NCE* (Percentile)	*Spring 1998 CTBS/4 Math Comp. NCE* (Percentile)	*Spring 1999 CTBS/4 Math Concepts NCE* (Percentile)	*Spring 1999 CTBS/4 Math Comp. NCE* (Percentile)	*Spring 2000 CTBS/5 Math Concepts NCE* (Percentile)	*Spring 2000 CTBS/5 Math Comp. NCE* (Percentile)
All BCP Schools (n=93)[24]	32.7 (21st)	41.8 (35th)	36.1 (25th)	36.1 (25th)	39.9 (32nd)	41.0 (34th)	45.8 (42nd)
All Control Schools (n=47)	34.9 (24th)	43.2 (37th)	41.3 (34th)	39.3 (31st)	34.7 (23rd)	45.0 (41st)	43.7 (38th)

Analyses of achievement test data indicate a striking impact of Direct Instruction on mathematics computation scores. Among the original kindergarten cohort, DI students moved, on average, from the 16th percentile at the end of first grade to the 48th percentile at the end of third grade (compared with growth among control counterparts from the 27th to 36th percentile over the same period). For the original kindergarten cohort, we examined the effect of three years (grades 1-3) of Direct Instruction on both math computation and math concepts scores, controlling for pretest scores on the Peabody Picture Vocabulary Test and demographic variables. Over that period, there is a highly significant effect of Direct Instruction on math computation scores ($p<.0005$, effect size=.52). There was no measurable effect on math concept scores ($p=.465$, effect size=.08). A total of 17% of the DI cohort was performing at grade level or above in math concepts, compared to 28% of control students. It is unclear, however, to what extent the DI students may have been lower than control students in mathematics readiness before Direct Instruction, as the first mathematics achievement scores available are at the end of one year of DI.[25]

Students who received three years of Direct Instruction in mathematics beginning in third grade[26] made greater NCE gains in math computation than their control counterparts and were nearing grade level (42nd percentile) by the end of fifth grade. Regression analyses (controlling for demographic factors and pretest scores) indicate a relatively strong effect of DI on math computation scores ($p<.0005$, effect size=.43) for this cohort over three years. Though DI students made steady gains in math concepts, these gains were not significantly larger than those of control students (when demographic factors and school readiness pretest measures are

[24] Fewer students also took the concepts subtests.

[25] Though among the six pairs of schools overall there is no significant difference between experimental and control schools in the kindergarten pretest measure (Peabody Picture Vocabulary Test), there is a significant difference, favoring control schools, when we include just the four pairs in which Direct Instruction in mathematics has continued uninterrupted for this cohort since the fall of 1997. Since there is a reasonably high correlation between PPVT score and mathematics achievement (though understandably lower than between PPVT and reading achievement measures), it is reasonable to assume that, overall, students at DI schools began math instruction with a significant disadvantage.

[26] This analysis includes just the three BCP schools in which DI math instruction began in third grade in the fall of 1997 (and their controls).

controlled), and the DI students still demonstrated lower achievement levels in math concepts, on average, than their control counterparts (34th percentile vs. 41st percentile).

Since sample sizes were so greatly reduced over the study, we also analyzed test score results for the full group of third and fifth graders at DI and control schools in 1999-2000, many of whom had not been at the school for the full four-year period. We conducted regression analyses, controlling for previous year's math score, demographic variables (race, sex, free-lunch status), attendance, and mobility (whether at the same school as the previous year). Though such an analysis is able to detect only a one-year effect of Direct Instruction on mathematics achievement, it provides useful information for urban school districts with particularly mobile student populations. The effect of one year of Direct Instruction on math computation scores was significant at the third grade level (p=.029, effect size=.19). Direct Instruction did not, however, have a positive one-year effect on math concepts scores at the third grade level, and had no one-year effect on either computation or concepts scores at the fifth grade level.

Tables 9 and 10 below show the sizes of the NCE gains over the year (not controlled for demographic differences and pre-existing achievement differences). Taken together with the multivariate regression analysis that controls for these differences, the results suggest that both DI and control schools are making positive progress in raising student mathematics achievement. Average computation achievement at DI schools is definitely rising, but on average, students at both DI and control schools remain below grade level in mathematics achievement.

Table 9. Mean CTBS Scores and Gain Scores in Math Computation and Concepts for Third Grade Cohort at BCP and Control Schools, Spring 1999–Spring 2000
(Not Necessarily Same School in Spring 1999 and Spring 2000)

	Spring 1999 (2nd Grade)		*Spring 2000* (3rd Grade)		*Comp. Gain*	*Concepts Gain*
School (Number of students)	*Comp. Mean NCE* (Standard Dev.)	*Concepts Mean NCE* (Standard Dev.)	*Comp. Mean NCE* (Standard Dev.)	*Concepts Mean NCE* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)
4 BCP Schools (n=174)[27]	37.7 (22.9)	32.9 (18.9)	45.0 (19.9)	35.5 (14.9)	7.3 (16.4)	2.6 (16.5)
4 Control Schools (n=163)	35.3 (20.7)	33.8 (20.2)	38.7 (17.5)	39.5 (18.6)	3.4 (18.4)	5.7 (18.6)

[27] N is number of students who took both computation tests, which differs slightly from the number of students taking both concepts tests.

Table 10. Mean CTBS Scores and Gain Scores in Math Computation and Concepts for Fifth Grade Cohort at BCP and Control Schools, Spring 1999–Spring 2000
(Not Necessarily Same School in Spring 1999 and Spring 2000)

	Spring 1999 (4th Grade)		*Spring 2000* (5th Grade)		*Comp. Gain*	*Concepts Gain*
School (Number of students)	*Comp. Mean NCE* (Standard Dev.)	*Concepts Mean NCE* (Standard Dev.)	*Comp. Mean NCE* (Standard Dev.)	*Concepts Mean NCE* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)	*Mean NCE Gain* (Standard Dev.)
3 BCP Schools (n=138)	38.2 (20.2)	34.7 (19.6)	44.6 (19.9)	40.2 (18.1)	6.4 (19.9)	5.5 (13.1)
3 Control Schools (n=114)	28.9 (18.3)	32.2 (21.9)	38.7 (20.2)	37.9 (18.1)	9.8 (16.2)	5.7 (13.2)

Maryland School Performance Assessment Program (MSPAP) Outcomes

The Maryland State Department of Education holds individual schools accountable for student performance primarily through the Maryland School Performance Assessment Program (MSPAP), which began in 1993. MSPAP was designed to measure "how well students relate and use knowledge from different subject areas and how well they apply what they have learned to solve real world problems." It assesses not only basic skills and knowledge (reading, writing, and mathematics skills) but also "higher order skills such as supporting an answer with information; predicting an outcome and comparing results to the prediction; and comparing and contrasting information" (Maryland State Department of Education, 1999; also see Yen & Ferrara, 1997). Testing occurs each year in May in grades 3, 5, and 8 in six subjects (reading, writing, language usage, mathematics, science, and social studies). In year 4 of this study, the original cohorts from the study schools were in grades 3 and 5 and participated in MSPAP testing.

Schools judged as not making significant progress on MSPAP are designated by the State Department of Education as "eligible for reconstitution," and required to submit to close monitoring by state officials of their school improvement plan and its implementation. Of the six original BCP schools, three were designated reconstitution-eligible in 1996, and a fourth in 1997. Two of these are paired with control schools that have also been named reconstitution-eligible, while the control schools for the other two reconstitution-eligible BCP schools have not been so designated.

Analysis of the impact of BCP (Direct Instruction and Core Knowledge curricula) on student achievement using MSPAP scores is problematic, because change over time is in school-level scores, not the more clearly relevant change in students. Because individual student scores are not yet available for MSPAP, we are not able to distinguish between students who have been in the BCP or control schools from the beginning of the implementation and those students new to the schools. This limitation requires us to assume that non-longitudinal students' parents chose to bring their children to the experimental (BCP) and control schools for reasons independent of

the ongoing BCP implementation. In this context, MSPAP becomes a conservative test of the effects of the reform. Presumably it would be more difficult to show any reform's effects on measures that include students who did not receive the full treatment.

Table 11 presents longitudinal data on the MSPAP composite index for each BCP school and its paired control school. Scores represent the percentage of students at the school scoring "satisfactory or above" on the test.[28] The scores from the 1996 MSPAP administration are used as a pre-BCP-implementation measure, and are compared with the 2000 (end of fourth year) test results to calculate a four-year gain score. BCP schools showed somewhat higher gains overall than their control schools (4.9 points compared to 2.3 points for the six pairs; 3.4 compared to 0.7 points for four pairs). On average, control schools had a higher composite index before the reform began in the fall of 1996, and still had a somewhat higher index in the spring of 2000.

Tables 12-15 summarize the mean change from 1996 to 2000 for BCP and control schools in the percentage of students at both the third and fifth grade levels (the cohorts of interest in year 4) scoring satisfactory or higher on the six MSPAP subtests. On average, MSPAP scores have risen by 3 to 5 points at BCP schools since 1996, the year before the BCP reform began. Over this four-year period at the fifth grade level, BCP schools have outgained their controls by more than 6 points (since control schools declined, on average, at the fifth grade level). Control schools just slightly outgained BCP schools (by less than one point, on average) at the third grade level over this period. This suggests that in MSPAP gains overall, BCP schools are doing as well or better than matched controls (though neither set of schools is doing as well as the citywide average gain).

Table 11. Composite Index Scores of Maryland School Performance Assessment Program for BCP and Control Schools, 1993-2000

School	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*Change ('93-'00)*	*Change ('96-'00)*
BCP School 1	28.4	23.6	38.5	38.6	42.6	43.4	50.1	51.5	23.1	12.9
Control School 1	25.9	30.6	28.7	32.7	20.6	34.5	28.7	22.8	-3.1	-9.9
BCP School 2	12.7	5.5	7.6	16.8	6.8	12.4	9.7	12.7	0.0	-4.1
Control School 2	18.1	20.8	38.7	40.8	40.5	38.3	46.1	46.8	28.7	6.0
BCP School 3	7.6	12.5	14.3	20.4	15.4	19.1	18.5	21.4	13.8	1.0
Control School 3	18.1	17.0	17.8	17.0	20.2	17.6	19.7	20.3	2.2	3.3
BCP School 4	15.5	5.1	10.6	8.6	11.0	6.3	14.6	10.4	-5.1	1.8
Control School 4	5.5	13.6	7.3	12.0	18.5	9.7	15.8	12.6	7.1	0.6
BCP School 5	14.9	13.1	10.6	13.7	10.1	9.3	22.1	17.0	2.1	3.3
Control School 5	8.0	11.7	16.9	20.7	10.6	16.8	30.4	41.2	33.2	20.5
BCP School 6	2.5	2.1	6.0	4.2	6.5	12.0	12.3	18.9	16.4	14.7
Control School 6	8.3	7.4	12.1	19.5	15.4	19.0	20.2	12.5	4.2	-7.0

[28] By comparison, statewide levels of "% satisfactory" in reading have ranged close to 40%.

Table 12. Mean Change from 1996 to 2000 in Percentages of Third-Grade Students Obtaining Scores of "Satisfactory" or Higher on the Six Subtests of MSPAP: Six BCP Schools and Six Control Schools Versus Baltimore City Averages

	Change from 1996 to 2000			*Change Difference in Schools in Study and All Baltimore Schools*		
Subtest	***All Baltimore Schools***	***6 Control Schools***	***6 BCP Schools***	***Control Gain Relative to All Baltimore***	***BCP Gain Relative to All Baltimore***	***BCP Gain Relative to Control***
Reading	+7.3	+2.0	+0.7	-5.3	-6.6	-1.3
Math[29]	+5.6	+3.8	-6.9	-1.8	-12.5	-10.7
Social Studies	+12.1	+11.2	+6.9	-0.9	-5.2	-4.3
Science	+7.4	+5.0	+3.4	-2.4	-4.0	-1.6
Writing	+10.7	+7.1	+14.4	-3.6	+3.7	+7.3
Language	+7.9	+5.8	+11.8	-2.1	+3.9	+6.0
6 Subtest Mean	+8.5	+5.8	+5.1	-2.7	-3.5	-0.8

Table 13. Mean Change from 1996 to 2000 in Percentages of Fifth-Grade Students Obtaining Scores of "Satisfactory" or Higher on the Six Subtests of MSPAP: Six BCP Schools and Six Control Schools Versus Baltimore City Averages

	Change from 1996 to 2000			*Change Difference in Schools in Study and All Baltimore Schools*		
Subtest	***All Baltimore Schools***	***6 Control Schools***	***6 BCP Schools***	***Control Gain Relative to All Baltimore***	***BCP Gain Relative to All Baltimore***	***BCP Gain Relative to Control***
Reading	+9.0	-1.4	+5.4	-10.4	-3.6	+6.8
Math	+7.3	-11.7	-1.6	-19.0	-8.9	+10.1
Social Studies	+6.0	-3.9	+3.3	-9.9	-2.7	+7.2
Science	+10.1	+1.8	+8.1	-8.3	-2.0	+6.3
Writing	+1.9	-8.4	-1.7	-10.3	-3.6	+6.7
Language	+11.0	+8.9	+11.8	-2.1	+0.8	+2.9
6 Subtest Mean	+7.6	-2.5	+4.2	-10.0	-3.3	+6.7

[29] Math scores in this table do not include pairs 1 & 5, since implementation of the program at these original BCP schools was not continuous. In the fifth grade table, pair 2 is also excluded, since implementation began a year later.

Table 14. Mean Change from 1996 to 2000 in Percentages of Third-Grade Students Obtaining Scores of "Satisfactory" or Higher on the Six Subtests of MSPAP: Four BCP Schools and Four Control Schools Versus Baltimore City Averages [30]

	Change from 1996 to 2000			*Change Difference in Schools in Study and All Baltimore Schools*		
Subtest	*All Baltimore Schools*	*4 Control Schools*	*4 BCP Schools*	*Control Gain Relative to All Baltimore*	*BCP Gain Relative to All Baltimore*	*BCP Gain Relative to Control*
Reading	+7.3	-1.4	+2.2	-8.7	-5.1	+3.6
Math	+5.6	+3.8	-6.9	-1.8	-12.5	-10.7
Social Studies	+12.1	+8.8	+3.5	-3.3	-8.6	-5.3
Science	+7.4	+3.4	+1.7	-4.0	-5.7	-1.7
Writing	+10.7	+5.7	+11.5	-5.0	+0.8	+5.8
Language	+7.9	+4.9	+11.0	-3.0	+3.1	+6.1
6 Subtest Mean	+8.5	+4.2	+3.8	-4.3	-4.7	-0.4

Table 15. Mean Change from 1996 to 2000 in Percentages of Fifth-Grade Students Obtaining Scores of "Satisfactory" or Higher on the Six Subtests of MSPAP: Four BCP Schools and Four Control Schools Versus Baltimore City Averages

	Change from 1996 to 2000			*Change Difference in Schools in Study and All Baltimore Schools*		
Subtest	*All Baltimore Schools*	*4 Control Schools*	*4 BCP Schools*	*Control Gain Relative to All Baltimore*	*BCP Gain Relative to All Baltimore*	*BCP Gain Relative to Control*
Reading	+9.0	+1.7	+2.1	-7.3	-6.9	+0.4
Math	+7.3	-11.7	-1.6	-19.3	-8.9	+10.4
Social Studies	+6.0	-5.0	+2.4	-11.0	-3.6	+7.4
Science	+10.1	-3.6	+9.9	-13.7	-0.2	+13.5
Writing	+1.9	-7.5	-1.8	-9.4	-3.7	+5.7
Language	+11.0	+6.5	+7.1	-4.5	-3.9	+0.6
6 Subtest Mean	+7.6	-3.3	+3.0	-10.9	-4.6	+6.3

BCP schools have posted particularly high gains on the language subtest for both third and fifth grades (11.8 points at each grade level), and on the writing subtest for third grade (14.4 points). These gains hold (at slightly reduced levels) even when only the four NIFDI schools are included in the analysis. At the fifth grade level between 1996 and 2000, the percentage of

[30] These tables exclude pairs 1 and 5, which pursued implementation of Direct Instruction differently than NIFDI directed, and thus did not technically remain part of BCP. BCP School 1 also has a much lower free lunch rate than the other schools. In the fifth grade table, pair 2 was also excluded from the math analysis, since implementation of the DI math program began later at this grade level in this school.

students at BCP schools scoring satisfactory or above declined only in math and writing (and this decline was less than that for control schools). Only in math at the third grade level did BCP schools lose ground (an average decline of 6.9 points) while their control schools gained (an average increase of 3.8 points).

Since most of the BCP schools and their controls have higher than average percentages of low-income students, it is not surprising that these schools post lower than average gain scores compared with the district as a whole. Similarly, one would not expect these schools to perform better than the district average on MSPAP, at least this early in the reform process. It is noteworthy, therefore, that all but one of the BCP schools scored above the district average on the third grade language subtest in the spring of 2000 (with the sixth school scoring within 0.2 points of the district average). And half of the BCP schools scored at or above the district average on the fifth grade subtests in science, writing, and language.

Overall, these data tend to contradict the concerns voiced by teachers regarding how well Direct Instruction prepares students for the MSPAP. When compared to similar types of schools, rather than to the district as a whole, BCP schools are generally making similar or greater gains (even though they still score below comparison schools, as they did before the reform). At the same time, we cannot ignore the declines in math scores at BCP schools, and the gap between BCP and control schools in this subject in third grade. This may be a temporary phenomenon related to implementation issues and grade level effects. It is important to note the smaller decline in BCP school math scores at the fifth grade level, and the fact that control schools lost more ground at this level and the gap was reversed. The significant positive effect of Direct Instruction on math computation achievement (as measured by CTBS scores, discussed previously in this report) cannot be ignored. But the relatively low performance of BCP schools on the MSPAP math test raises important questions about how well DI math instruction prepares students for how they will need to use math in the future.

DISCUSSION AND CONCLUSION

Our analysis of implementation issues indicated that the marriage between Direct Instruction and Core Knowledge in the Baltimore Curriculum Project was heavily dominated by Direct Instruction during its first four years, and there was little evidence of integration of the two reforms. Implementation rates of the Direct Instruction component were relatively high, while, by comparison, implementation of the Core Knowledge component was much lower. Given the requirements for Direct Instruction, as defined by the developers, it is not yet clear to us whether Core Knowledge will ever be more than an additional component (as contrasted with an integral part) of the reform effort, though BCP staff members were pursuing greater integration of the reforms as this study concluded. Our analysis of student outcomes was an evaluation primarily of Direct Instruction, because its implementation levels were considerably higher than those of the Core Knowledge component.

The evidence presented above indicates that student outcomes improved in BCP schools. The evidence is mixed on whether outcomes were significantly better for BCP students than for control students. Perhaps most striking was the lower rate of grade retention in BCP cohorts

compared to control cohorts. Direct Instruction was also clearly more effective than the control curricula in producing higher achievement in mathematics computation. There was also short-term evidence of a positive impact on reading vocabulary test scores and measures of oral reading fluency, but no compelling evidence in the first four years of an impact on reading comprehension scores and mathematics concepts and applications (the primary dependent variables specified in the original evaluation plan). Though growth in reading comprehension and mathematics concepts achievement occurred for students receiving Direct Instruction, that growth was not significantly greater than for students receiving other types of instruction. At the four high-poverty schools in the study, student achievement remained below grade in reading and mathematics even after students received several years of DI instruction.

Those who emphasize the detrimental effects of retention in grade (e.g., Natriello, 1998; Owings & Magliaro, 1998) would laud the low retention rate achieved at BCP schools compared with their control school counterparts. At first glance, the low retention rates at BCP schools compared with control schools were impressive. On the other hand, there was no evidence that BCP students achieved significantly higher in reading comprehension than their control counterparts, and lower achievers who were not retained at BCP schools were often reading from stories at a lower grade level, often in reading groups with children at a lower grade level. While these BCP children did not endure the negative social consequences of formal retention, they still experienced some of the effects of retention (grouping with younger children, learning opportunities pitched at the lower grade level). Only further longitudinal analyses will determine whether there is a long-term advantage to the form of social promotion practiced in BCP schools. There may indeed be a cost savings, if BCP students finish school without the cost of an additional (retention) year.

In conclusion, we interpret the findings presented in this report as evidence that Direct Instruction is a viable whole-school reform option for raising student achievement in reading and mathematics, if implemented at the same levels as in this study. Though DI may not necessarily perform better than other curricular alternatives, it produced sufficient achievement gains to justify its continuation as a reform option. In schools where teachers have become heavily invested in the program and scores are rising, we believe it is particularly important to continue implementing the reform, because change could be potentially disruptive. Based on the evidence from this four-year study, we would recommend that schools consider Direct Instruction as one of several reform options aimed at boosting student achievement, and make their choices based on the needs of their students and the capacities and preferences of their teaching staffs.

REFERENCES

Adams, Gary L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.

Berends, M., Kirby, S. N., Naftel, S., & McKelvy, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up*. Washington, DC: Rand Corporation.

Berkeley, M. (2002). The importance and difficulty of disciplined adherence to the educational reform model. *Journal of Education for Students Placed At Risk, 7*, 221-239.

Borman, G., Hewes, G., Rachuba, L., & Brown, S. (2002). *Comprehensive school reform and achievement: A meta-analysis* (Report 59). Baltimore, MD and Washington, DC: Center for Research on the Education of Students Placed At Risk.

Brent, G., & DiObilda, N. (1993). Effects of curriculum alignment versus Direct Instruction on urban children. *Journal of Educational Research, 86*, 333-338 .

Brown, J.S., Collins, A., Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Core Knowledge Foundation (1995). *Core Knowledge sequence*. Charlottesville, VA: Core Knowledge Foundation.

CTB (1991). *CTBS/4 technical report*. Monterey, CA: Macmillan/McGraw Hill.

CTB/McGraw-Hill (1997). *TerraNova technical bulletin 1*. Monterey, CA: CTB/McGraw Hill.

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 512-538). New York: Longman.

Datnow, A., McHugh, B., Stringfield, S., Hackler, D.J. (1998). Scaling up the Core Knowledge sequence. *Education and Urban Society, 30* (3), 409-432.

Datnow, A., Borman, G., & Stringfield, S. (2000). School reform through a highly specified curriculum: Implementation and effects of the Core Knowledge sequence. *Elementary School Journal, 101*(2), 167-191.

Davis, G. (1999, October 27). Transcripts of nonpublished interview transcript conducted by Martha Mac Iver and Elizabeth Kemper on October 27, 1999 in Baltimore, MD as part of the qualitative research underlying this article.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test–Revised.* Circle Pines, MN: American Guidance Service.

Engelmann, Z., & Madigan, K. (1996). *Direct Instruction coaching manual.* Eugene, OR: Engelmann-Becker Corporation.

Everston, C.M., & Burry, J. (1989). Capturing classroom context: The observation system as lens for assessment. *Journal of Personnel Evaluation in Education, 2,* 297-320.

Fuchs, D., & Deno, S. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232-243.

Hall, T. E., & Tindal, G. (1989). Using curriculum-based measures to group students in reading. In G. Tindal, K. Essick, C. Skeen, N. George, & M. George (Eds.), *The Oregon Conference Monograph, 1989* (pp. 10-12). Eugene, OR: University of Oregon, College of Education.

Hasbrouck, J., & Tindal, G. (1991). Curriculum-based oral reading fluency norms for students grades 2-5. *Teaching Exceptional Children, 24,* (3) 41-44.

Herman, R., Aladjem, D., McMahon, P., Massem, E., Mulligan, I., O'Malley, A., Quinones, S., Reeve, A., & Woodruff, D. (1999). *An educator's guide to schoolwide reform.* Washington, DC: American Institutes for Research.

Hirsch, E.D., Jr. (1988). *Cultural literacy.* New York: Vintage.

Hirsch, E.D., Jr. (1996). *The schools we need.* New York: Doubleday.

Kearnes, D. & Anderson, J. (1996). Sharing the vision: Creating New American Schools. In S. Stringfield, S. Ross, & L. Smith (Eds.), *Bold plans for school restructuring* (pp. 9-23). Mahwah, NJ: Erlbaum.

Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed At Risk, 1* (2), 147-169.

Lambert, C., & Reynolds, J.E. (1997). *The state of Baltimore's schools: Data and information on the status of Baltimore City Public Schools. (Vols. 1-3).* Baltimore, MD: Advocates for Children and Youth.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry.* Sage: Newbury Park, CA.

Mac Iver, M., Kemper, E., & Stringfield, S. (2000). *The Baltimore Curriculum Project: Third year report.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Mac Iver, M., McHugh, B. & Stringfield, S. (2000). *Implementation and effects of the Maryland Core Knowledge Project: Final (fifth year) evaluation report* (Report 50). Baltimore, MD and Washington, DC: Center for Research on the Education of Students Placed At Risk.

Mac Iver, M., McHugh, B., & Stringfield, S. (1999). *The Baltimore Curriculum Project: Second year report.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Mac Iver, M., Stringfield, S., & Hall, T. (1999). *Final report on the use of curriculum-based measures in the Baltimore Curriculum Project.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Marston, D., & Deno, S. (1982). *Implementation of direct and repeated measurement in the school setting* (Research Report No. 106). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Marston, D., Deno, S., & Tindal, G. (1983). *Comparison of standardized achievement tests and direct measurement tests in measuring pupil performance* (Research Report No. 126). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Maryland State Department of Education (1999). *Maryland school performance report: 1999.* Baltimore, MD: Author.

McHugh, B., & Spath, S. (1997). Carter G. Woodson Elementary School: The success of a private school curriculum in an urban public school. *Journal of Education for Students Placed at Risk, 2* (2), 121-135.

Natriello, G. (1998). Failing grades for retention. *School Administrator, 55*(7), 14-17.

Nunnery, J. A. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society, 30* (3), 277-295.

Owings, W.A., & Magliaro, S. (1998). Grade retention: A history of failure. *Educational Leadership, 56,* 86-88.

Patton, M.Q. (1990). *Qualitative evaluation and research methods.* Newbury Park, CA: Sage.

Rosenshine, B. (2002). Helping students from low-income homes read at grade level. *Journal of Education for Students Placed At Risk, 7* (2), 273-283.

Schaffer, E., & Nesselrodt, P. (1993). *Special Strategies Observation System.* Charlotte, NC: University of North Carolina-Charlotte.

Slavin, R. (1987). A theory of school and classroom organization. *Educational Psychologist, 22,* (2), 89-108.

Stallings, J. (1980). Allocated learning time revisited, or beyond time on task. *Educational Psychologist, 22,* 89-108.

Stanovich, K. (1991). Word recognition: Changing perspectives. In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 418-452). New York: Longman.

Stringfield, S. (1995). *Fourth-year evaluation of the Calvert Program at Barclay Elementary School.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Stringfield, S., Millsap, M., Winfield, L., Brigham, N., Yoder, N., Moss, M., Nesselrodt, P., Schaffer, E., Bedinger, S., & Gamse, B. (1997). *Urban and suburban/rural special strategies for educating disadvantaged children: Second year report.* Washington, DC: U.S. Department of Education.

Stringfield, S., McHugh, B., & Datnow, A. (1998). *The Baltimore Curriculum Project: First year report.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Thrift, G. (2000, August 2). Transcripts of nonpublished interview transcript conducted by Martha Mac Iver and Elizabeth Kemper on August 2, 2000 in Baltimore, MD as part of the qualitative research underlying this report.

Viadero, D. (2001). Whole-school projects show mixed results. *Education Week, 21*(10), pp. 1, 24-45.

Yen, Y., & Ferrara, S. (1997). The Maryland School Performance Assessment Program: Performance assessment with psychometric quality suitable for high-stakes usage. *Educational and Psychological Assessment, 57*(1), 60-88.

APPENDIX

Comparison of BCP and Control Schools

Though we attempted to select schools with similar demographic characteristics as matched controls for the BCP schools, demographics have changed somewhat over time (Table A1) and cohorts under study at each school in year 4 do have significant differences (Table A2). Though the school selected as a control school for BCP School 2 had similar published aggregate free lunch rates in 1995-96, the published rates have fluctuated and there is now a 20 percentage point differential favoring the control school. Although this difference is unfortunate for our study, data on a school readiness measure (Peabody Picture Vocabulary Test) shows no significant difference between the entering kindergarten cohorts of these matched schools, in either 1996-97 or 1999-2000. In pairs 3 and 4, there are also proportionately more students eligible for free lunch at the BCP school. On the other hand, there are also two pairs of schools where the BCP school now has considerably fewer students eligible for free lunch than does the control school (pairs 1 and 5). Because of this difference on an important demographic variable, we conducted analyses controlling for all individual background characteristics (sex, race, and free-lunch status).

What evidence do we have that students at BCP and paired control schools were achieving at basically the same level before the introduction of the BCP intervention? Table A3 reports mean percentile scores on the Peabody Picture Vocabulary Test, administered to BCP kindergarteners in November 1996, and to control students generally later in the year. While the different administration dates may present some problems in comparing results, overall the groups do not differ significantly.[31] Unfortunately, it was not possible to administer a CTBS/4 pretest to the 1996-97 second grade control cohort, so we can't be certain of the comparability of these students with the corresponding BCP cohort. Multivariate analyses are able, however, to control for spring 1997 achievement levels in examining differences between BCP and control students on spring 1998 achievement tests, which allows us to test for the effect of the past year of BCP instruction.

Table 11 (in the main body of this report) has composite MSPAP scores over time for the six BCP schools and their control schools. Though these scores tend to be quite volatile, it does appear that at least two of the control schools (#2 and #6) consistently scored more than five points higher than their paired BCP school in the years before the BCP initiative (1993-96). This may indicate some previous advantage that might affect the results of outcomes analyses. We seek, however, to adjust for these pre-existing differences by controlling for demographic and pretest variables in all analyses.

---

[31] Students at BCP School 1 score significantly better than those at the control school. Responding to the concern that administering the PPVT in November, two months after the beginning of the BCP intervention in classrooms, was not a true pretest of student academic ability, researchers administered the test to the 1997-98 kindergarten cohort at one of the BCP schools and found no significant difference between the mean student score in September and November. Unfortunately, however, students at the four schools that continued to implement the full Direct Instruction reform under NIFDI scored significantly lower as a group on the PPVT than students at the four control schools. For this reason, we control for the PPVT score in all multivariate analyses.

Table A.1. Background Characteristics of BCP and Control Schools

School	*Free Lunch*					*Special Ed*					*Entrants*				
	1996	*1997*	*1998*	*1999*	*2000*	*1996*	*1997*	*1998*	*1999*	*2000*	*1996*	*1997*	*1998*	*1999*	*2000*
BCP School 1	39.2	34.5	33.8	33.6	22.1	7.8	7.7	8.5	9.1	6.3	6.4	6.1	5.4	10.1	11.4
Control School 1	50.5	51.9	52.8	61.8	68.1	14.6	14.5	17.9	15.8	15.4	12.8	12.8	12.4	22.0	21.5
BCP School 2	95.3	92.5	95.6	95.6	89.6	12.4	11.2	10.5	13.8	12.7	17.7	12.6	15.0	12.8	22.3
Control School 2	95.8	72.8	94.3	94.1	62.9	13.4	12.5	12.8	13.1	12.4	16.1	20.6	9.7	13.0	27.9
BCP School 3	79.9	76.5	68.2	70.7	78.3	12.2	11.1	11.4	13.5	12.0	9.9	9.7	9.8	15.7	17.2
Control School 3	60.3	58.8	61.4	63.6	63.8	14.8	16.7	16.2	13.1	12.7	16.9	13.9	15.8	34.7	30.8
BCP School 4	84.4	89.3	89.7	87.3	83.0	9.2	13.7	12.8	10.2	14.0	14.4	21.3	14.2	31.5	37.7
Control School 4	60.3	75.8	64.8	72.1	66.5	12.2	22.0	21.5	14.9	11.2	19.4	13.5	22.1	29.9	10.7
BCP School 5	72.9	79.9	85.7	84.1	66.3	19.4	18.1	15.8	16.4	16.5	11.2	10.8	13.8	10.6	NA
Control School 5	90.0	85.7	91.7	86.0	86.7	22.7	19.1	16.5	20.6	15.6	11.6	14.5	16.8	13.8	23.9
BCP School 6	97.9	90.0	93.9	88.2	91.9	12.5	15.0	15.3	13.6	14.2	10.7	15.7	16.4	20.8	25.5
Control School 6	91.4	81.9	93.4	73.4	93.8	10.2	11.7	15.1	15.3	13.2	11.0	15.6	7.7	20.1	31.8

Table A.2. Background Characteristics of BCP and Control School Student Cohorts, 1999-2000

School	*% Male*		*% Free Lunch*		*% African-American*	
	3rd	*5th*	*3rd*	*5th*	*3rd*	*5th*
BCP School 1	55.8	58.8	22.1	27.5	55.8	60.0
Control School 1	49.5	42.9	65.9	62.2	79.1	82.7
BCP School 2	30.9	42.2	85.3	88.9	100.0	100.0
Control School 2	47.0	40.0	65.2	50.0	100.0	98.0
BCP School 3	51.2	56.1	68.3	66.3	8.5	4.1
Control School 3	58.0	58.2	66.7	68.7	17.3	13.4
BCP School 4	45.2	54.8	81.0	67.7	19.0	29.0
Control School 4	60.0	43.3	80.0	63.3	48.9	30.0
BCP School 5	45.1	48.0	57.7	56.0	100.0	100.0
Control School 5	42.4	37.9	84.8	96.6	100.0	100.0
BCP School 6	44.6	55.2	87.5	96.6	96.4	98.3
Control School 6	50.7	50.7	91.5	92.4	98.6	97.5
Total BCP Schools	46.0	53.2	64.4	63.6	63.4	61.5
Total Control Schools	51.7	46.7	73.9	70.0	71.8	71.8

Table A.3. Mean Peabody Picture Vocabulary Test Percentiles[32] for Kindergartners at BCP and Control Schools

School	*Mean Percentile for All Kindergartners*	*Mean Percentile for All Kindergartners Who Continued in First Grade at the Same School*	*Date*
BCP School 1	47	47	10/30/96
Control School 1	18	21	11/18/96
BCP School 2	8	8	11/4/96
Control School 2	8	9	1/14/97
BCP School 3	15	17	11/15/96
Control School 3	20	24	12/16/96
BCP School 4	14	13	11/6/96
Control School 4	19	18	3/21/97
BCP School 5	11	16	11/7/96
Control School 5	12	14	3/12/97
BCP School 6	4	4	11/7/96
Control School 6	9	9	5/9/97
All BCP Schools	14	15	*
All Control Schools	14	14	*

[32] Average percentiles were calculated after averaging NCE scores.

CRESPAR

Johns Hopkins University
Center for Social Organization of Schools
3003 North Charles Street – Suite 200
Baltimore MD 21218
410-516-8800 / 410-516-8890 fax

Howard University
2900 Van Ness Street, NW
Washington DC 20008
202-806-8484 / 202-806-8498 fax

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").